# UC Santa Cruz
## UC Santa Cruz Previously Published Works

**Title**

Integrated Proteogenomic Characterization across Major Histological Types of Pediatric Brain Cancer.

**Permalink**

https://escholarship.org/uc/item/6fn4631t

**Journal**

Cell, 183(7)

**Authors**

Petralia, Francesca

Tignor, Nicole

Reva, Boris

et al.

**Publication Date**

2020-12-23

**DOI**

10.1016/j.cell.2020.10.044

Peer reviewed

# Integrated Proteogenomic Characterization across Major Histological Types of Pediatric Brain Cancer

*A full list of authors and affiliations appears at the end of the article.*

## SUMMARY

We report a comprehensive proteogenomic analysis, including whole genome sequencing, RNA sequencing, proteomic and phosphoproteomic profiling, of 218 tumors across 7 histologic types of childhood brain cancer: low grade glioma (n=93), ependymoma (32), high grade glioma (25), medulloblastoma (22), ganglioglioma (18), craniopharyngioma (16) and atypical teratoid rhabdoid

*Correspondence: RESNICK@email.chop.edu (A.C.R.), storm@email.chop.edu (P.B.S.), BROOD@childrensnational.org (B.R.R.), pei.wang@mssm.edu (P.W.).

SECONDARY AUTHORS

Alicia Francis, Allison M. Morgan, Angela J. Waanders, Angela N. Viaene, Anna Maria Buccoliero, Arul M. Chinnaiyan, Carina A. Leonard, Cassie N. Kline, Chiara Caporalini, Christopher R. Kinsinger, Chunde Li, David E. Kram, Derek Hanson, Elizabeth Appert, Emily A. Kawaler, Eric H. Raabe, Eric M. Jackson, Jeffrey P. Greenfield, Gabrielle S. Stone, Gad Getz, Gerald Grant, Guo Ci Teo, Ian F. Pollack, Jason E. Cain, Jessica B. Foster, Joanna J. Phillips, July E. Palma, Karen A. Ketchum, Kelly V. Ruggles, Lili Blumenberg, Macintosh Cornwell, Mahdi Sarmady, Marcin J. Domagalski, Marcin P. Cie lik, Mariarita Santi, Marilyn M. Li, Matthew J. Ellis, Matthew A. Wyczalkowski, Meghan Connors, Mirko Scagnet, Nalin Gupta, Nathan J. Edwards, Nicholas A. Vitanza, Olena M. Vaske, Oren Becher, Peter B. McGarvey, Ron Firestein, Sabine Mueller, Samuel G. Winebrake, Saravana Mohan Dhanasekaran, Shuang Cai, Sonia Partap, Tatiana Patton, Toan Le, Travis D. Lorentzen, Wenke Liu, William E. Bocik.

Consortia

Clinical Proteomic Tumor Analysis Consortium (CPTAC)

Alexey I Nesvizhskii, Alicia Francis, Amanda G Paulovich, Ana I Robles, Anna Calinawan, Antonio Iavarone, Antonio Colaprico, Arul M Chinnaiyan, Azra Krek, Bing Zhang, Boris Reva, Christopher R Kinsinger, D. R Mani, David Fenyö, Dmitry Rykunov, Emily Boja, Emily A. Kawaler, Eric E Schadt, Felipe da Veiga Leprevost, Francesca Petralia, Gad Getz, Gonzalo Lopez, Guo Ci Teo, Henry Rodriguez, Hui-Yin Chang, Jacob J Kennedy, Jamie Moon, Jeffrey R Whiteaker, Jiayi Ji, John Spytz, Joshua Wang, Jun Zhu, Karen A Ketchum, Karin D Rodland, Karl K Weitz, Karsten Krug, Kelly V Ruggles, Lei Zhao, Li Ding, Liang-Bo Wang, Lili Blumenberg, Lizabeth Katsnelson, Maciej Wiznerowicz, MacIntosh Cornwell, Marcin J Domagalski, Marcin P Cie lik, Marina A Gritsenko, Matilda Broberg, Matthew J Ellis, Matthew A. Wyczalkowski, Matthew E Monroe, Mehdi Mesri, Nathan J Edwards, Nicole Tignor, Pei Wang, Peter B McGarvey, Richard G Ivey, Richard D Smith, Ronald J Moore, Rosalie K Chu, Rui Zhao, Sanjukta Guha Thakurta, Saravana Mohan Dhanasekaran, Selim Kalayci, Seungyeul Yoo, Shrabanti Chowdhury, Shuang Cai, Stephan Schürer, Steven P Gygi, Tao Liu, Tara Hiltke, Tatiana Omelchenko, Toan Le, Travis D Lorentzen, Uliana J Voytovich, Vasileios Stathias, Vladislav A Petyuk, Weiping Ma, Wenke Liu, William E Bocik, Xi Steven Chen, Xiaoyu Song, Ying Wang, Zeynep Gümü .

Children's Brain Tumor Tissue Consortium (CBTTC)

Adam C Resnick, Nithin D. Adappa, Elizabeth Appert, Valerie Baubet, Oren Becher, Miguel A. Brown, Anna Maria Buccoliero, Jason E. Cain, Chiara Caporalini, Meghan Connors, Brian M. Ennis, Bailey Farrow, Ron Firestein, Jessica B. Foster, Krutika S. Gaonkar, Gerald Grant, Jeffrey P. Greenfield, Yiran Guo, Nalin Gupta, Derek Hanson, Allison P. Heath, Eric M. Jackson, Cassie N. Kline, Mateusz Koptyra, David E. Kram, Sarah Leary, Carina A. Leonard, Chunde Li, Marilyn M. Li, Jena V. Lilly, Robert M. Lober, Jennifer L. Mason, Allison M. Morgan, Sabine Mueller, Javad Nazarian, Judy E. Palma, James Palmer, Sonia Partap, Tatiana Patton, Joanna J. Phillips, Ian F. Pollack, Eric H. Raabe, Pichai Raman, Samuel Rivero-Hinojosa, Jo Lynne Rokita, Brian R. Rood, Mariarita Santi, Mahdi Sarmady, Mirko Scagnet, Gabrielle S. Stone, Phillip B. Storm, Lamiya Tauhid, Olena M. Vaske, Angela N. Viaene, Nicholas A. Vitanza, Angela J. Waanders, Samuel G. Winebrake, Nathan Young, Bo Zhang, Yuankun Zhu.

AUTHOR CONTRIBUTIONS

Study Conception & Design: H.R., A.C.R., A.G.P., S.P.G., P.B.S., B.R.R., P.W.; Performed Experiment and data collection: M.K., Y.Z., A.S.H., J.L.R., J.M., J.V.L., B.F., NY, A.C.R., S.P.G., S.G.T., J.S., A.G.P., T.D.L., U.J.V., L.Z., J.J.K., R.G.I., J.R.W., M.A.G., K.K.W., J.M., V.A.P., R.K.C., R.J.M., M.E.M., R.Z., R.D.S., K.D.R., T.L.; Data Analysis, M.A.B., Y.G., K.S.G., P.R., B.M.E., B.Z., Y.Z., F.D.L., H.C., A.I.N., F.P., A.I., N.T., S.C., W.M., J.J., X.S., B.R.,D.R.,A.K., P.W., G.L.,S.Y., J.Z, V.S., S.S., A.C., X.S.C., B.R., A.C,R., D.F., M.B., J.W., J.M.W., Y.W., K.V.R., L.B., M.C., W.L.,E.K., L.K.; Sample Resources: A.C.R., J.V.L., J.M., P.B.S., V.B., L.T.; Writing: F.P, N.T., M.K., S.C., W.M., M.A.B., K.S.G., J.L.R., D.R., A.K., X.S., J.J., S.Y., A.C.R B.R.R., P.W. ; Visualization, S. K., G.L, Z.G, A.P.C., F.P., D.R., A.K., N.T.,W.M., S.C., X.S., J.J.; Supervision, P.W., B.R.R., A.C.R., A.G.P., S.P.G., J.L.R., Y.Z., M.K., A.S.H., P.B.S., D. F.; Project Administration, P.W., H.R., E.B., T.H., C.R.K., M.M., F.P., N.T., B.R.R., A.C.R. ; Funding Acquisition, P.W., B.R.R., P.B.S., A.C.R., A.G.P., S.P.G., T.L., R.D.S., D.F., A.I.N., L.D., D.R.M., E.E.S., B.Z., H.R.; All authors contributed to data interpretation, manuscript editing and revision.

**DECLARATION OF INTERESTS** Dr. Eric Schadt serves as Chief Executive Officer for Sema4 and has an equity interest in this company.

tumor (12). Proteomic data identifies common biological themes that span histologic boundaries, suggesting that treatments used for one histologic type may be applied effectively to other tumors sharing similar proteomic features. Immune landscape characterization reveals diverse tumor microenvironments across and within diagnoses. Proteomic data further reveal functional impacts of somatic mutations and CNVs not evident in transcriptomic data. Kinase-substrate association and co-expression network analysis identifies important biological mechanisms of tumorigenesis. This is the first large-scale proteogenomic analysis across traditional histologic boundaries to uncover foundational pediatric brain tumor biology and inform rational treatment selection.

## INTRODUCTION

Pediatric brain tumors are the leading cause of cancer related deaths in children (Ostrom et al., 2018). While genomic techniques have begun to illuminate the pathogenesis of many pediatric brain tumors, there are some unique challenges that limit the translation of these findings into new effective therapies. Since pediatric brain tumors have a relatively low mutational burden (Chalmers et al., 2017; Grobner et al., 2018; Northcott et al., 2017; Parsons et al., 2011; Pugh et al., 2012; Robinson et al., 2012), the majority of pediatric brain tumors defy treatment approaches that exploit targetable genomic events. In addition, many pediatric brain tumors are characterized by aberrant epigenetic landscapes, but there is as yet no effective way to specifically target these key programmatic changes (Capper et al., 2018). RNA profiling has identified subgroups within histologic diagnoses and highlighted pathways thought to be active in these groups. But targeting these pathways has largely been unsuccessful. Potential explanation for this lack of translation is that these mechanisms reside many regulatory layers away from the primary functional element of the cell, the protein (Rivero-Hinojosa, 2018).

In recent years, quantitative mass spectrometry and bioinformatic analyses have matured, resulting in the ability to add a quantitative proteomic facet to a primarily genomic-based biological understanding of diseases (Clark et al., 2019; Dou et al., 2020; Gillette et al., 2020; Mertins et al., 2016; Rivero-Hinojosa et al., 2018; Zhang et al., 2014; Zhang et al., 2016). These efforts have shown a distinct uncoupling of RNA transcript abundance from protein abundance, particularly in cancer. This fact alone could account for a significant disconnect between genome-based biological discovery and clinical validation. Analysis of these integrated proteogenomic data sets has the potential to aid in the identification of new therapeutic avenues.

Another challenge of translating new molecular findings into therapeutic innovations is that the subdivision of traditional histology-based entities into molecular subgroups fragments patient populations into ever smaller groups creating existential challenges for clinical trial design. This is especially true for rare cancers such as pediatric brain tumors. One salient feature of proteomics is the ability to discern biology closer to cellular intent by virtue of its focus on the main functional moiety of the cell, the protein. Because disparate upstream genomic events can result in similar downstream pathways and patterns, by scrutinizing these resultant events, proteomics can identify common biology across histologic and molecular boundaries.

In an attempt to incorporate proteomics into a biological understanding of pediatric brain tumors, we undertook the first large-scale comprehensive proteogenomic analysis inclusive of the genomics, transcriptomics, global and phosphoproteomics of a large cohort of 218 tumor samples representing 7 distinct histologic diagnoses, including low grade glioma (LGG), ependymoma(EP), high grade glioma (HGG), medulloblastoma (MB), ganglioglioma, craniopharyngioma(CP) and atypical teratoid rhabdoid tumor (ATRT). Unsupervised clustering based upon the proteome revealed surprising alignments between subsets of tumor diagnoses previously regarded as biologically distinct and led to a number of insights herein described. We seek to demonstrate that the incorporation of the proteomic and phosphoproteomic dimensions into this large-scale multi-omic study leads to functional insight that will help drive translational efforts.

## RESULTS

### Proteogenomic analyses of pediatric brain tumor specimens

For 218 fresh frozen tumor samples from 199 patients representing 7 histologic types of pediatric brain tumors, we performed whole genome sequencing (WGS), RNA sequencing (RNAseq), quantitative proteomic and phosphoproteomic profiling. All samples were sourced from Children's Hospital of Philadelphia. Figure 1A illustrates the sample distribution across 7 histologic types: LGG (n=93), EP (32), HGG (25), MB (22), ganglioglioma (18), CP (16) and ATRT (12).

For proteomic and phosphoproteomic quantitation, all 218 tissue samples were analyzed by liquid chromatography and triple mass spectrometry with tandem-mass-tag (TMT) isobaric labeling. The number of proteins and phosphosites measured per sample ranged from 4661 to 5731 (median 5122) and 2155 to 3415 (median 2714) respectively. In total, we identified and quantified 8802 proteins and 18,235 phosphosites. Among them, 6429 proteins and 4548 phosphosites were observed in more than 50% of the samples of at least one histologic diagnosis and were considered in the downstream analysis. In addition, 440 phosphosites from ischemia-induced proteins (Mertins et al., 2014) were excluded to avoid any artificial effect induced by variations in sample collection.

WGS and RNAseq were also performed for most samples. After quality filtering, somatic mutation, DNA copy number alterations and RNAseq based gene expression data were derived for 200, 190 and 188 tumor samples, respectively (Fig. 1B, Star Method). All processed proteogenomic data sets can be queried, visualized and downloaded from http://pbt.cptac-data-view.org/.

### Proteogenomic clustering of pediatric brain tumors

Consensus clustering based on global proteomic data identified eight clusters (Fig. S1A) with distinct survival outcomes (Fig. 1C), stemness scores, proliferation indices and pathway activities (Fig. 1D and Table S1). We termed the eight clusters: *Ependy, Medullo, Aggressive, Cranio/LGG-BRAF$^{V600E}$, HGG-rich, Ganglio-rich, LGG BRAF$^{WT}$-rich, and LGG BRAF$^{Fusion}$-rich*.

While some clusters coincided with histologic diagnoses, such as *Medullo*, other clusters contained a mixture of different diagnoses (Figs. S1B). Firstly, the *Cranio/LGG BRAF^{V600E}* cluster *(C4)* aligned a subset of CP tumors with LGG tumors harboring *BRAF^{V600E}* mutations, while the rest CP were aligned with the *LGG BRAF^{WT}-rich* cluster *(C8)* (Figs. 1D, S1A). This segregation of the CP samples into two distinct clusters was also supported by parallel consensus clustering analysis based on phosphorylation data (Figs. S1A). Division of CP samples, however, was not detected based on RNA data (Fig. S1A), and the sample-wise correlation between proteomic and RNAseq profiles was rather low for samples in these two protein clusters (Fig. 1D).

While *CTNNB1* mutation is an important oncogenic factor for pediatric CP (Campanini et al., 2010), the proteomic clusters, C4 and C8, did not distinguish *CTNNB1* mutation status (Fig. 1E). Instead, they more closely resembled the patterns induced by *BRAF^{V600E}* in LGG (Figs. 1D, 5A). *BRAF^{V600E}* mutation, an oncogenic event for some adult CP (Brastianos et al., 2014), has not been previously detected in pediatric CP patients. Our findings suggest that a subset of pediatric CP tumors, despite their lack of *BRAF^{V600E}* mutations, showed similar proteomic changes as those in *BRAF^{V600E}* LGG tumors. This motivates the hypothesis that some pediatric CP might benefit from MEK inhibitor (MEKi) based treatment, a strategy that has been used for *BRAF^{V600E}* LGG tumors (Fangusaro et al., 2019) and has shown preclinical promise in adult CP (Apps et al., 2018). Indeed, a group of genes suggested to be downregulated by MEKi (Pratilas et al., 2009) were found to be upregulated in the CP samples from C4 (Fig. S1C). Furthermore, downstream proteins/ substrates of MEK/ERK kinases, including ERK1/2, were upregulated in these samples (Figs. 1E, 1F, S1C), a known consequence of *BRAF^{V600E}* mutation.

In addition, central members of the AKT/mTOR pathway also showed higher kinase activity in C4 compared to C8 CP samples (Figs. 1E, 1F, S1C), consistent with the contrast between *BRAF^{V600E}* and *BRAF^{WT}* LGG tumors (Fig. 5B). The AKT pathway has been implicated as a resistance pathway emerging after RAF/ERK inhibition in BRAF driven tumors (Jain et al., 2017). Preclinical studies have demonstrated the value of coordinated inhibition of MEK and mTOR, the primary AKT effector, in LGG (Jain et al., 2017). Our findings further suggest the potential application of this rationale for some of the CP patients.

Note, upregulation of key MEK/ERK/AKT kinases in C4 compared to C8 are only visible based on kinase activity assessment using phosphoproteomic data but are not reflected in RNA/protein abundance, suggesting the important complementary role of phosphoproteomic data (Fig. S1C).

To validate TMT measurements of proteins and phosphosites of interest, targeted Mass Spectrometry experiments of a customized protein/phosphoprotein marker panel were applied to the same set of tumor samples following immuno-Multiple-Reaction-Monitoring (MRM) experiment protocols (Whiteaker et al., 2018). The MRM measurements of key players in the MEK/ERK pathways confirm the substantial differences in C4 and C8 CP (Fig. 1G). Moreover, these MRM assays can accurately classify the two subtypes of CP as reflected by their high AUC values (Fig. 1G), suggesting the feasibility of classifying these subtypes in clinical practice by using MRM-based assays.

Another proteomic cluster containing a mixture of diagnoses is the *Aggressive* cluster, characterized by poor survival outcomes (Fig. 1C). EP in the *Aggressive* cluster were more similar to tumors within cluster, regardless of histology, than to the other EP in the *Ependy* cluster (Fig. S1A). Specifically, members belonging to the evolutionarily conserved multifunctional polymerase-associated factor 1 complex (PAF1C), including PAF1, CDC73, CTR9, LEO1 and RTF1, were found to be significantly upregulated in the *Aggressive* cluster compared to the *Ependy* cluster (Fig. S1D, S1E). PAF1C plays a vital role in gene regulation and has been implicated in tumorigenesis (Moniaux et al., 2006; Tomson and Arndt, 2013). PAF1C regulates a variety of factors involved in histone covalent modifications, transcription, and mRNA 3' end processing (Karmakar et al., 2018) (Fig. S1D). These factors all showed upregulation patterns in the *Aggressive* cluster compared to the *Ependy* cluster based on global and phosphoproteomic data (Fig. S1E). Neither the segregation of EP into different clusters nor consistent upregulation of PAF1C members and downstream players were observed in RNA data (Figs. S1A, S1E).

Note, while the 9 samples from the post-mortem collection (Fig. 1B) blended well with other surgically obtained samples in the protein/RNA based clustering results, they grouped together in one phosphoproteomic cluster (Fig. S1A), suggesting caution when studying PTM based on post-mortem samples. To avoid any potential artificial effects, the post-mortem samples were not considered in the downstream analyses involving phosphoproteomic data.

## Immune infiltration in pediatric brain tumor

We performed cell type deconvolution analysis using xCell (Aran et al., 2017) based on RNA data to infer relative abundance of different cell types in the tumor microenvironment (Fig. 2A, Table S2). The inferred proportion of neuronal and microglia cells were further confirmed based on signatures derived from single-cell RNAseq study of glioblastoma (Fig. 2A) (Darmanis et al., 2017). Consensus clustering based on inferred cell proportion identified five sets of tumors with distinct immune and stromal features: *Cold-medullo*, *Cold-mixed*, *Neuronal*, *Epithelial* and *Hot* (Fig. 2A). Comparing this to proteomic clusters, we observed lower immune infiltration in more aggressive proteomic clusters such as *Aggressive*, *Medullo* and *Ependy*, while higher immune infiltration in *LGG BRAF$^{WT}$-rich*, *LGG BRAF$^{Fusion}$-rich* and *Cranio/LGG BRAF$^{V600E}$* (Figs. 2D, S2A).

The *Hot* group, containing a mixture of LGG, HGG and ganglioglioma samples, was characterized by the presence of multiple types of immune cells, including macrophages, microglia and dendritic cells (Fig. 2A). As expected, compared to other tumors, the *Hot* cluster showed the upregulation of immune related pathways including epithelial mesenchymal transition (EMT) (Lou et al., 2016) (Figs. 2A, 2B). Moreover, adenosine producers (e.g., phosphatases ENTPD1 and NT5E), which have been shown to protect against inflammatory oxidative stress, inhibit immune activators and activate immunosuppressing cells (Chisci et al., 2017; Kordass et al., 2018), were upregulated based on both RNA and protein data in the *Hot* cluster (Figs. 2A, 2B), suggesting adenosine reducing therapies can be investigated for these tumors (Lakka and Rao, 2008; Leone and Emens, 2018; Perrot et al., 2019).

*Neuronal* also contained a mixture of LGG, HGG and Ganglioglioma tumors, but was uniquely characterized by the upregulation of Glutamate Receptor Signaling and Neurotransmitter Transport pathways involved in neuron communication and activation of cell growth (Pereira et al., 2017; Stepulak et al., 2014)(Figs. 2A, 2B and Table S2). This observation may support a glutamate/glutamate-receptor mediated mechanism of glioma progression in these samples. Recent studies demonstrate that glutamatergic synapses exist between neurons and glioma cells in pediatric (Venkatesh et al., 2019) and adult (Venkataramani et al., 2019) HGG. Glutamate receptors can activate Ca2+/calmodulin dependent protein kinase II (CAMK2A/B/G/D), which engages PI3-kinase (*PIK3CA*) and signals to RAS through centaurin-α1 (*ADAP1*) (Hayashi et al., 2006). Consistently, we found that *GRIA1*, *CAMK2A/B/G/D*, *PIK3CA* and *ADAP1* all showed significant upregulation in *Neuronal* (Table S2), further suggesting the active role of glutamate signaling in the *Neuronal* group. In addition, high levels of glutamate can promote immune evasion mechanisms (Cai et al., 2018), and indeed we observed decreased gene expression of *CD4*, *CD8A* and macrophage related genes in *Neuronal* cluster as compared to the *Hot* cluster (Table S2). At the same time, the *Neuronal* cluster was characterized by upregulation of pathways of energy metabolism such as OXPHOS, Mitochondrial Protein Complex and Glycolysis solely based on proteomic data (Figs. 2A–C, S2C and Table S2). It has been reported that glutamine blockade induces divergent metabolic programs to overcome tumor immune evasion, and glutamine antagonism could serve as a "metabolic checkpoint" for tumor immunotherapy (Leone et al., 2019), which might benefit tumors like the ones in the *Neuronal* cluster.

LGG tumors, which were split into the *Neuronal* and *Hot* clusters, showed substantial tumor microenvironment heterogeneity (Fig. 2A). Interestingly, $BRAF^{V600E}$ and $BRAF^{Fusion}$ events, important oncogenic drivers of LGG tumors, showed significant association with multiple immune signatures. In particular, APM class I genes were upregulated in both $BRAF^{Fusion}$ and $BRAF^{V600E}$ tumors compared to wild type (Figs. 2E, 2F and Table S2). More careful investigation of pro-inflammatory (M1) and pro-regenerative (M2) macrophage and microglia signatures (Fig. S2B and Table S2) based on markers specific to these cell types (Dello Russo et al., 2017; Fumagalli et al., 2018; Krasemann et al., 2017) further suggests that M1 macrophages and M2 microglia were upregulated in $BRAF^{Fusion}$ compared to wild type (Fig. 2E). The significant difference between microglia and macrophage polarization across BRAF statuses is further illustrated in Fig. 2G: $BRAF^{Fusion}$ promoted more M2 microglia, while $BRAF^{V600E}$ promoted more M2 macrophages. This observation is in concordance with the balance between macrophage and microglia polarization reported for adult glioblastoma (Darmanis et al., 2017).

The *Epithelial* cluster, containing as expected only CP tumors which originate from odontogenic epithelium, was characterized by the upregulation of EMT, immune related pathways as well as CTLA4 and PD-1 molecules (Figs. 2A, 2B and Table S2). Therefore, CP could potentially benefit from immune checkpoint therapy as previously reported (Coy et al., 2018).

Finally, both *Cold-medullo* and *Cold-mixed* exhibited upregulation of Signaling by WNT, Beta Catenin TCF Complex Assembly, Regulation of Apoptosis and Proteasome. This is

consistent with the recent reports that tumors with active WNT Signaling were characterized by lower levels of immune infiltration (Luke et al., 2019). Again, these patterns of upregulation were observed in both *Cold-medullo* and *Cold-mixed* clusters based on proteomic and phosphoproteomic data but not RNA data (Figs. 2A, 2B, 2C, S2C).

## Integrative proteogenomic analyses reveal functional consequences of mutation and CNV

While pediatric tumors usually have fewer genetic alterations compared to adult tumors (Grobner et al., 2018), a few recurrent DNA alterations were observed in this cohort (Fig. S3A). We first evaluated the impact of the few somatic mutations on the corresponding RNA/protein levels. LGG tumors with *BRAF^V600E^* mutation had significantly downregulated BRAF protein abundance compared to *BRAF^WT^* LGG tumors (Fig. 3A), while the reduction was not significant at the transcript level (Fig. S3B). *CTNNB1* mutation resulted in elevated protein/RNA levels among CP samples, while *NF1* mutation resulted in the downregulation of cognate protein and transcript in HGG (Figs. 3A, S3B). *SMARCB1* RNA/protein were significantly downregulated in ATRT samples compared to other diagnoses as expected, and the downregulation was the result of different types of DNA alterations, including mutation, deletion, and copy neutral LOH (Fig. S3C).

In terms of genomic instability, MB, HGG and EP tumors showed relatively higher genomic instability (Fig. S3A). By integrating copy number, RNA and proteomic data, we detected 1,541 genes whose transcript and protein abundance were simultaneously influenced by their own CNVs in one or more diagnosis, referred to as CNV-RNA/Protein cis-cascade events (Fig. S3D and Table S3). In addition, for 515 of these 1,541 genes, we detected significant dependence between their phosphosite abundance and CNV in one or more diagnoses, referred to as CNV-RNA/Protein/Phospho cis-cascade events (Fig. S3D and Table S3). These lists of cis-cascade events facilitate the identification of important players in frequently amplified/deleted genome regions. One example is *RABGAP1L (1q25)*, an EP CNV-RNA/Protein/Phospho cis-cascade gene (Figs. 3B, 3C), whose amplification is associated with GTPase activation and RAB-GTPase binding (Itoh et al., 2006) and has been reported to be an independent predictor of tumor progression in EP (Kilday et al., 2012). Another member from the *RAB GTPase* gene family, *RAB3GAP2 (1q41)*, which has a key role in neurodevelopment (Ng and Tang, 2008), was identified to be a CNV-RNA/Protein cis-cascade gene for EP, MB and HGG tumors. Our analysis further pinpoints an important player in maintenance of glioblastoma stemness, *FDPS*, proximate to *RABGAP1L*, as a CNV-RNA/Protein cis-cascade gene for HGG (Abate et al., 2017; Kim et al., 2018a).

While recurrent amplification of *RABGAP1L*, *RAB3GAP2* and *FDPS* were observed in all EP, MB, and HGG tumors, a significant influence of *RABGAP1L* amplification on its protein/phosphoprotein was only observed in EP, while *FDPS* was found to be a CNV/RNA/ protein cascade event only in HGG (Figs. 3B, 3C). On the other hand, for MB, only *RAB3GAP2* is identified as a CNV-RNA/Protein cis-cascade gene. These observations suggest that CNV of the same genomic region could lead to different functional perturbations in different diagnoses.

We then studied the *trans*-regulatory effects of somatic mutations and CNVs on proteins and phosphoproteins within each diagnosis. Besides *BRAF* mutation/fusion in LGG (discussed

below), the only other profound *trans*-regulatory effects were detected between mutation of *CTNNB1*, which codes β-catenin, and many proteins and phosphosites in CP (Fig. 3D and Table S3). β-catenin is crucial for two developmental processes: establishment and maintenance of cell-type-specific cell-to-cell adhesion and regulation of target gene expression via the WNT Signaling pathway (Gao et al., 2018). As expected, *CTNNB1* mutation, which boosted β-catenin abundance, is found to be associated with upregulation of proteins/phosphosites related to cell-to-cell adhesion, as well as upregulation of members of the WNT Signaling pathway such as APC, GSK3A and GSK3B (Fig. 3D, 3E). Specifically, while phosphosite abundance of APC at *Ser 2812* showed significant elevation in *CTNNB1* mutation cases, this upregulation was not observed based on protein abundance of *APC* (Fig. 3E). It is well known that WNT signaling results in the liberation of β-catenin and its translocation to the nucleus where it binds to transcription factor (TCF) complexes to activate transcription (Fig. 3F). In our data, we observed significant association between RNA and protein/phosphosite abundance *of TCF4, TCF25* and *CTNNB1* mutation in CP. The interaction of β-catenin with *TCF4* has been proposed as a target for the development of anti-cancer drugs in other tumor types (Fasolini et al., 2003). However, we observed that both RNA expression and phosphosite abundance of *TCF4* were significantly lower in the *CTNNB1* mutated group. Instead, both RNA and proteomic abundance of *TCF25*, another transcription factor that may play a role in cell death control (Cai et al., 2006), were upregulated in *CTNNB1* mutated CP. These results suggest that, among this group of CP, downstream effects of mutation in *CTNNB1* could be mediated by *TCF25*.

## Phosphoproteomic analysis of kinase activity

Because of the tremendous appeal of kinases as drug targets, it is of great importance to characterize the common and differential kinase activations within and across histologies. CDK1 and CDK2, essential cyclin-dependent kinases promoting the G2–M transition and regulating G1 progression and the G1–S transition (Santamaria et al., 2007), were elevated in more proliferative tumors including ATRT, MB, HGG and EP based on global abundance and kinase activity (Figs. 4A, S4B and Table S4), the latter derived from the abundance of phosphorylated substrates (StarMethod). The activation of CDK1 and CDK2 in more proliferative tumors was also confirmed by higher correlation between their global abundance and kinase activity scores (Fig. 4A). To further characterize the dependence of individual substrates on kinases, we constructed diagnosis-specific kinase-substrate networks leveraging experimentally validated kinase-substrate regulation database (Hornbeck et al., 2015) (StarMethod). Some kinase-substrate associations of CDK1/2 were shared across different diagnosis (Fig. S4A). For example, the association between CDK2 and MCM2 at Ser 139 was detected in ATRT/MB, EP, HGG and LGG. On the other hand, some Diagnosis-specific associations were detected, such as CDK2 and NPM1 at Ser 70 in HGG and LGG, and CDK2 and TERF2IP at Ser 203 in EP (Fig. S4A). All MCM2, NPM1 and TERF2IP (*RAP1*) play important roles in cell proliferation (Box et al., 2016; Fei and Xu, 2018; Schmitt and Stork, 2001), implying diverse mechanisms used by CDK2 to influence cell proliferation in various diagnoses.

Another important kinase is CAMK2A (Calcium/Calmodulin Dependent Protein Kinase II Alpha), which is directly involved in metastatic invasion of glioma cells (Chen et al., 2011;

Cuddapah and Sontheimer, 2010; Shin et al., 2019). While CAMK2A was the most abundant in ganglioglioma, a higher correlation between its kinase activity score and protein abundance was observed in HGG (Fig. 4A and Table S4). The inferred kinase-substrate network of HGG further highlights association between CAMK2A and GJA1 (connexin 43) at both Ser 325 and Ser 314 (Figs. 4B and S4A). Phosphorylation of connexin 43 at Ser 325 and Ser 314 promotes gap junction assembly between glioma and astrocytes (Cooper and Lampe, 2002) and drives cancer cell migration as well as glioma invasion (Behrens et al., 2010; Hong et al., 2015). Thus, our data suggests a potential role of CAMK2A in glioma invasion. Moreover, in HGG, CAMK2A protein abundance was found to be associated with SYN1 Ser 568 and SYN1 Ser 605 (Figs. 4B and S4A), the latter increases synaptic transmission and regulates synaptic vesicle dynamics (Magupalli et al., 2013). This further aligns with the relevant role of CAMK2A in glioma invasion, as glioma cells form functionally active synapses with neurons and neural activity mediated by neuron-to-glioma synapses drives glioma invasion and growth (Venkataramani et al., 2019; Venkatesh et al., 2019). Interestingly, the activation of CDK1/2 and CAMK2A, reflected by their elevated protein abundance respectively, tended to be exclusive of each other, suggesting the existence of two different signaling mechanisms among HGG tumors (Fig. 4C).

To further confirm the kinase activity of CDK2 and CAMK2A in HGG, we carried out independent TMT proteomic and phosphoproteomic experiments in an independent cohort of 23 pediatric and young adult HGG (StarMethod) and validated the kinase-phosphosite associations between aforementioned pairs (Fig. 4B). The negative correlations between CDK1/2 and CAMK2A protein abundance were also confirmed in this validation cohort (Fig. 4C), suggesting two different signaling mechanisms among HGG tumors.

Another interesting group of kinases, CDK5 and GSK3B, were upregulated in ganglioglioma and a subset of LGG belonging to the *Ganglio-rich* cluster (Fig. 4A). CDK5 and GSK3B have been suggested to be regulators of synapse formation, neurogenesis and cell proliferation (Cole, 2012; Shah and Lahiri, 2017). The kinase-phospho network revealed interesting associations between CDK5/GSK3B and their substrates in LGG (Figs. 4D, S4A), such as the phosphosites of ADD2 (beta-adducin). ADD2 is highly expressed in brain regions associated with high plasticity (e.g., hippocampus), involved in neuronal morphology, and required for synaptogenesis (Bednarek and Caroni, 2011; Porro et al., 2010). Positive associations between CDK5 and ADD2 at Ser 604, as well as GSK3B and ADD2 at Ser 693, reflected CDK5-dependent priming of GSK3B activity (Farghaian et al., 2011). Moreover, it has been shown that CDK5 regulates recruitment of SYN1 to nascent synapses (Easley-Neal et al., 2013), and phosphorylation of SYN1 by CDK5 at Ser 553 controls efficiency of neurotransmitter release (Qiao et al., 2014). Thus, the observed positive association between CDK5 and SYN1 Ser 553 in LGG might support an increase in synaptogenesis between glioma cells and neurons in LGG, which however needs to be further confirmed through functional studies. This increase also aligns with the association between CAMK2A1 and SYN1 at Ser 605 and Ser 568 (Figs. 4D, S4A), as phosphorylation of SYN1 at Ser 605 by CAMKII was shown to increase synaptic transmission (Magupalli et al., 2013). Furthermore, the positive association between CDK5 and STMN1 at Ser 38 highlights the importance of STMN1-mediated synaptogenesis for pediatric gliomagenesis, as stathmin phosphorylation, including Ser 38, is essential for synaptic plasticity and

memory, and increases synaptic strength through promoting microtubule stability and dendritic transport of the GluA2 subunit of AMPA-type glutamate receptors to the synapse (Uchida et al., 2014). All these observations are in line with the findings that gliomas can hijack neuronal development by creating neuron–glioma synapses (Venkataramani et al., 2019; Venkatesh et al., 2019), and link to the immune clustering results: the global abundance of CDK5 and GSK3B was upregulated in the subset of LGGs from the *Neuronal* immune cluster (Fig. S4C)

### Insights from proteogenomic analysis of LGG

In order to help discern biological insights stemming from the frequent targetable alterations of *BRAF* in LGG, we identified proteins associated with $BRAF^{V600E}$ mutation and $BRAF^{Fusion}$ (Table S5). Compared to $BRAF^{WT}$ tumors, $BRAF^{V600E}$ and $BRAF^{Fusion}$ cases showed both common as well as alteration-type specific changes (Fig. 5A). Particularly, in $BRAF^{V600E}$, we observed significant abundance changes of protein in the *MAPK* (ERK) signaling pathway (Fig. 5A) compared to $BRAF^{WT}$ tumors. *MAPKs* are the terminus of the RAS/RAF/MAPK pathway, inhibitors of which have been used to treat *BRAF* altered tumors of multiple cancer types, including brain tumors (Schreck et al., 2019). For instance, MEK inhibitor monotherapy recently showed promising results in low-grade pediatric glioma with *BRAF* alterations (Fangusaro et al., 2019). Investigation of an RNA expression based "MEK inhibitor signature" (Pratilas et al., 2009) in our data confirmed that genes downstream of MEK kinases are greatly upregulated in $BRAF^{V600E}$ as compared to the $BRAF^{WT}$ tumors (Fig. S5B), supporting the current usage of MEK inhibitor therapy for these LGG tumors. Moreover, RNA/protein abundance of AKT Serine/Threonine kinases AKT1 and AKT2, as well as RNA of AKT1S1 (Fig. 5B and Table S5) showed significant upregulation in $BRAF^{V600E}$ tumors. MRM experiments measuring AKT isoforms on same set of tumors further validated this upregulation (Fig. 5B). Indeed, the AKT pathway has been implicated as a resistance pathway emerging after RAF/MAPK inhibition in *BRAF* driven tumors via upregulation of receptor tyrosine kinases (Jain et al., 2017). Preclinical studies have demonstrated the value of coordinated inhibition of MEK and mTOR, the primary AKT effector, in LGG (Jain et al., 2017). Our findings further strengthen this rationale for upcoming clinical trials.

Next, we employed a network-based approach to study the impact of BRAF alterations on the phosphoproteome. Co-expression network analysis resulted in 18 closely connected modules capturing the association across phosphosites (Table S5). Interestingly, two modules (Figs. 5C, S5C) are significantly upregulated in $BRAF^{V600E}$ and $BRAF^{Fusion}$ samples, respectively (Table S5). Module 1 was significantly enriched in phosphosites associated with MYC Targets (Fig. 5C) and G2M Checkpoint, confirming the upregulation of Cell-Cycle related pathways in $BRAF^{V600E}$ compared to $BRAF^{WT}$ LGG patients (Fig. 5A). Moreover, Module 1 was significantly enriched in phosphosites regulated by AKT2 (Fig. 5D). Specifically, it contained phosphosites of a group of Heterogeneous Nuclear Ribonucleoproteins, including *HNRNPUL1* and *HNRNPUL2 (*Fig. 5C, 5D). Active AKT2 was reported to suppress the interaction between HNRNPU and caspase-9b, causing inhibition of apoptosis (Vu et al., 2013). Taken together, these observations further support the concept of inhibiting mTOR/AKT in $BRAF^{V600E}$ LGG.

On the other hand, Module 2 appeared to capture a group of phosphosites perturbed in *BRAF^Fusion^* but not in *BRAF^V600E^* cases (Fig. S5C–E). The top druggable kinase associated with these phosphosites in this module is PDGFRA, which encodes a cell surface tyrosine kinase receptor (Fig. S5D). PDGFRA is frequently mutated/amplified in pediatric HGG, and has been suggested to serve as a treatment target for pediatric HGG (Koschmann et al., 2016). Our data reveals a strikingly similar upregulation of PDGFRA protein/RNA in *BRAF^Fusion^* samples as that in HGG tumors (Fig. S5A), suggesting the exploration of *PDGFRA* targeted treatment in *BRAF^Fusion^* tumors as well.

### Insights from proteogenomic analysis of HGG

Isocitrate dehydrogenases (IDHs) are enzymes that catalyze the oxidative decarboxylation of isocitrate, producing α-ketoglutarate (KG) and CO2. Mutations of IDH1 and IDH2 proteins, which can be effectively targeted by drugs, have been found in ~80% of grade II and III astrocytomas, oligodendrogliomas, and secondary glioblastomas (Bergaggio and Piva, 2019). These mutations however are infrequent in pediatric HGG (~11%) (Kim and Liau, 2012), as also observed in our data. On the other hand, recent literature has reported prognostic and/or therapeutic roles for the wild type IDH genes/proteins in various adult cancers such as melanoma, glioblastoma and kidney (Bergaggio and Piva, 2019, Tanaka et al., 2013, Calvert et al., 2017), bringing interest to understand their roles in pediatric HGG tumors.

We first investigated associations between IDH proteins and overall survival (OS) of HGG patients. Since point mutations in histone *H3.3* (*H3F3A, H3K27M*) has been reported to lead to a worse prognosis in HGG (Karremann et al., 2018), *H3* status was adjusted for when assessing association between OS and abundance of IDH proteins (StarMethod). Strikingly, all IDH proteins showed positive association with improved OS among the *H3^WT^* group (Figs. S6A, 6A and Table S6). A parallel analysis based on RNA data detected similar associations between expression of *IDH1/2/3A* with OS (Fig. S6B). Consistently, the Oxidative-phosphorylation pathway, harboring IDH1/2/3, is one of the leading pathways whose up-regulation was significantly associated with improved OS among *H3^WT^* patients (Fig S6C).

While all IDH proteins showed positive correlation with OS, no correlation was observed between IDH1 and IDH2/3 protein abundances (Fig. 6B). Although this is not surprising, as IDH1 is situated in the cytosol and peroxisomes, whereas IDH2/3 are in the mitochondria, it implies potentially complementary information in IDH1 and IDH2/3 for prognostic prediction. Indeed, compensatory functions between IDH1 and IDH2 have been reported in acute myeloid leukemia (Zhang et al., 2019) and colorectal cancer (Koseki et al., 2015). These observations motivated us to evaluate the joint prognostic value of IDH proteins. In addition, since IDH3 proteins are highly correlated with IDH2, and IDH3 proteins are of relatively lower abundance compared to IDH1/2, we decided to focus on IDH1 and IDH2 to avoid collinearity in the analysis. Specifically, by jointly modeling IDH1 and IDH2 proteins in one multivariate Cox regression model, we estimated that, among *H3^WT^* HGG patients, the risk of death has a 23.58-fold increase, with a 95% confidence interval (CI) of [1.42, 384.6] (Fig. 6C), if the combined abundance of IDH1 and IDH2 is 50% lower (i.e., decrease

of 1 in the weighted log2 abundance, StarMethod). The extremely wide CI for the hazard ratio of IDH1/2 score is a result of the limited sample size in the analysis (n=19). To further verify this finding, we performed TMT proteomics profiling experiments for an additional 41 pediatric HGG samples including 23 from an independent patient cohort, and 18 from the existing study cohort with remaining tumor material (StarMethod). With this second dataset, we confirmed the association between the reduced expression of combined IDH1/2 protein abundances and shorter OS after accounting for additional confounders such as tumor location (Fig. 6D, 6F, S6E, StarMethod).

Unlike $H3^{WT}$ HGG tumors, IDH1/2 proteins showed an adverse effect on OS among $H3^{MUT}$ tumors (Figs. 6C, 6D). But due to the small number of the $H3^{MUT}$ tumors (n=7 and 12 in the discovery and second data sets respectively), further verification is warranted.

While the factors driving IDH1/2 protein abundances in these pediatric HGG tumors remain largely unknown, one possible factor is revealed by the CNV-RNA/Protein cis-regulation investigation, which identified IDH1 as a CNV-RNA/Protein cascade protein (Fig. S3D). As illustrated in Fig. S6D, *IDH1* deletion, which was observed in about 20% of the HGG tumors, significantly downregulated the protein abundance of IDH1.

Moreover, to nominate potential drug targets for pediatric HGG based on new insights from proteogenomic data, we performed a drug connectivity analysis to identify drug candidates whose impact on the transcriptome and proteome is diametrically opposed to the characteristics identified as central to HGG biology. Because of the lack of adjacent normal tissue of HGG patients, we chose to derive RNA/proteomic signatures of HGG aggressiveness by comparing HGG and LGG tumors. We then leveraged the LINCS L1000 transcriptomic and P100 phosphoproteomic perturbation-response databases to search for candidate drugs inducing effects that oppose the corresponding input (Litichevskiy et al., 2018; Subramanian et al., 2017) (StarMethod, Table S6). CDK inhibitors were predicted to reverse the aggressiveness of HGG based on both RNA and phosphoproteomic data (Figs. 6G, S6F/G, and Table S6). Consistently, the kinase activity of CDK1 and CDK2 were upregulated in HGG compared to LGG (Fig. 6H). MEK, proteasome and HDAC inhibitors were found to be significant based on RNA data alone (Figs. 6G, S6F, S6G). Although MEK substrates were not observed in the phosphoproteomic data of HGG samples, *MAPK1* kinase activity, downstream of MEK, was found upregulated in HGG tumors (Fig. 6H). The upregulation of substrates of CDK and MAPK1 proteins in HGG tumors supports that CDK and MEK inhibitors might be effective for HGG tumors.

## Comparison between initial and recurrent tumors

Earlier work has reported distinct patterns between initial and recurrent tumors of the same patient (Morrissy et al., 2016). Based on proteomic and genomic profiles of 18 pairs of surgical samples from two distinct disease occurrences of the same patients in our cohort, we tried to address the question of whether the recurrent tumors should be considered as independent tumors during treatment evaluation.

In all 18 pairs, the recurrent/progressive (RP) disease was of the same histologic diagnosis as the initial (IN) tumor. RP tumors carried 0%-52% (mean 18%) of the IN tumor mutations

(Fig. 7A), which was lower than that of adult GBM (Cancer Genome Atlas Research, 2008) and LGG (Cancer Genome Atlas Research et al., 2015) (Fig. S7A). Remarkably, all three MB progression samples had a *TP53* mutation that was absent in their paired primary tumors (Fig. 7A), consistent with the observation by Morrissy et al., 2016. In addition, there was an increase in chromosome arm aberrations in the RP samples, with the number of breakpoints increasing on average from 32 to 53 (Fig. 7A). In contrast, in adult GBM and LGG tumors, CNV events from primary tumors were similar as that in their recurrences (Fig. S7A). Proteomic profiles also revealed differences between RP and IN tumors. In fact, most primary and recurrent samples were classified into different proteomic clusters (Fig. 7A). As an example, one *BRAF$^{WT}$* LGG case (pair 173.2154), with BRAF being wild type in both its IN and RP tumors, had its IN tumor allocated to the *LGG BRAF$^{WT}$-rich* proteomic cluster and the RP tumor allocated to the *Cranio/LGG BRAF$^{V600E}$* cluster. Consistent with the characteristics of these two proteomic clusters, we observed the upregulation of RNA Transcription and splicing/EMT and Coagulation and downregulation of Gap junction in the RP tumors as compared to their IN counterparts (Figs. 1D, 7A, and Table S7). Consistently with the allocation of RP tumor to the Cranio/LGG BRAFV600E cluster, a higher kinase activity for ERK1/ERK2 was observed in RP tumor compared to IN tumor (Fig. 7C), suggesting that a MEK inhibitor therapy might be more beneficial for the RP tumor. Another LGG case (pair 350.944) had the IN tumor allocated to *Cranio/LGG BRAF$^{V600E}$*, while the RP sample allocated to the *LGG BRAF$^{WT}$-rich* cluster, which resulted in opposite trend in pathway activities followed by a reversed trend in the activity of ERK1/ERK2 (Figs. 7A, 7C). These changes in pathway activation highlight the need for de novo characterization of recurrent cases which might impact on treatment decisions.

We also investigated the correlation between IN and RP proteome profile pairs (Figs. 7B, S7B). The fact that a good number of primary-recurrent pairs were not highly correlated with each other at the proteomic level supports the idea that recurrence (or progression) of a tumor could have different tumorigenesis mechanisms. Based upon these observations, an approach that assesses the molecular properties of recurrent events independent of the initial tumor seems to be warranted.

## DISCUSSION

This study represents the very first attempt to perform a large scale of proteogenomic integrative analysis for multiple distinct pediatric brain tumor diagnoses in an effort to discover new effective targeted therapies. High quality genomic, transcriptomic, proteomic and phosphoproteomic data were generated as a public resource from a retrospective cohort of 218 frozen tissue samples collected by a single institution.

Although histologic diagnosis remains the cornerstone of classifying tumors into therapeutic categories, it is now well recognized that molecular subgroups within histologically similar tumors can be identified on the basis of transcriptomics, genomics and methylomics. Our study is based on the recognition that proteomics/phosphoproteomics needs to be integrated with other omics to gain an improved system biology view of molecular subgroups. In addition, we advocate the importance of characterizing biological themes that cross histologic boundaries and unite individual tumors of disparate histologies and cells of origin,

because such insights can lead to new extension of treatment shown to be effective in one type of tumor to be used on other histologically disparate tumors sharing the same proteomic features. For example, our proteomic/phosphoproteomic data clustering analyses revealed two distinct subgroups of pediatric CP, with one subgroup showing strikingly similar proteomic/phosphoproteomic characteristics as pediatric LGG $BRAF^{V600E}$ tumors. This observation suggests the potential use of MEK/MAPK inhibitors in a subset of pediatric CP, which currently has no robust chemotherapeutic options.

The existence of two subgroups of pediatric CP, however, is not evident from RNA data. Similarly, we observed profound discordance between RNA and protein abundance in other histologies, such as EP and LGG. The low/moderate RNA-protein correlation observed in this project is consistent with other large scale proteogenomic projects (Clark et al., 2019; Dou et al., 2020; Gillette et al., 2020). Multiple factors, such as protein turnover and selective translation, contribute to the low correlation between RNA and protein abundance (Clark et al., 2019; Dou et al., 2020). Interestingly, more aggressive tumors tend to show increased protein-RNA correlation, a phenomenon observed across multiple cancer proteogenomic studies (Clark et al., 2019; Dou et al., 2020). One possible explanation is that aggressive tumors often have high proliferation and the boosted translation activities in highly proliferative (tumor) cells result in more correlated RNA and protein signals. Thus, studying the proteome reveals insights not evident from RNA-based analysis alone.

There is also significant value in the integration of large scale proteomic and genomic data to identify the ramifications of genomic events on biological function. A good illustration of this contribution is the ability to discern at the protein level the cis effects of copy number alterations by tracking the cascade of abundance from gene dose to transcript level to protein/phosphosite abundance. In this way, the relevant genes in a chromosomal region with altered copy number can be identified for validation of their biologic contribution (e.g., *RAGBAP1L* in EP and *FDPS* in HGG with *1q* gain).

It must be noted that a clear eyed view of the proteomic contribution needs to acknowledge a presumed equivalence between abundance and activity. However, the investigation of kinase activity based on phosphoproteomics showed that protein abundance can be reduced in active signaling pathways (e.g., ERK1/3 in CP), which could happen due to feedback loops in many complicated regulatory processes, suggesting the important role of phosphoproteomics data for characterizing pathway activities.

Many molecularly targeted agents have now been sufficiently characterized in terms of their safety and mechanisms of action to allow for combinations of these agents to enter clinical trials. The selective targeting of multiple kinase nodes within these networks represent a common strategy to construct more effective treatments while leaving fewer redundant escape routes for the tumor cell. Concurrently targeting the MEK and mTOR pathways in RAF activated tumors is a strategy that has emerged from investigations of resistance pathways resulting from RAF/MAPK inhibition. Our characterization of proteome and phosphoproteome changes due to *BRAF* aberrations lends further rationale to this approach in LGG where resistance per se is not an issue but durability of response after treatment cessation is.

The retrospective study design enabled us to access follow up clinical data, including outcomes. This strategy also allowed us to study rare diagnoses and compare primary and recurrent tumors from the same patients. One interesting finding was that the protein abundances of the wild type IDH1/2 are prognostic in HGG tumors without H3K27M mutation. In addition, by comparing primary and recurrent specimens, we were able to discern significant differences between these paired specimens. The shifts in underlying biology accompanying tumor recurrence necessitate independent assessment and therapeutic decisions for those recurrent tumors.

This project represents a significant advance in the biological interrogation of pediatric brain tumors at multiple levels of biological control and across traditional histologic boundaries. While the limited sample sizes of some histologies pose a significant limitation on certain investigations, as the first proteogenomic characterization of such histologies, the data and analytical results from this project serve as a valuable resource. Importantly, it is the result of a necessarily expansive partnership between children's hospitals; their patients and families; philanthropic and federal funding; and physician scientists and computational biologists. Such endeavors demonstrate the potential of large scale proteogenomic science and the power of inclusive collaboration to tackle a pervasive threat to our children, pediatric brain tumors.

### Limitations and Future Directions

While this project represents the most comprehensive multi-omic analysis of pediatric brain tumors ever undertaken, there are nevertheless a number of limitations that are the result of the rarity of the tumor types studied and the nature of the samples available. 1) There are additional layers of cellular regulation not included in this initiative such as methylation profiling, histone mark profiling, ribo-seq, metabolomics, and acetyl-proteomics. 2) This study was structured within tissue access limitations to provide the ability to discern common biology across major histologic types of pediatric brain cancer. In so doing, we sacrificed the ability to perform in depth proteomic analyses within the tumor types that were represented by smaller sample sizes. A future study gathering larger cohorts of the less common brain tumor types would be instructive in identifying the biology that is unique to those tumors. 3) This study leveraged retrospective tissue collection which allowed for the analysis of survival outcome and initial/progressive tumor pairs as well as making it feasible to study rare tumors. The cost of this approach however was that the samples would not be fully exploited for phospho-proteomics and our phosphosite data, while significant and high quality considering the amount of tissue available, would have been deeper if we had samples that had been prospectively acquired with a phospho-proteomics specific collection protocol.

While acknowledging these limitations, we can also see some tantalizing future possibilities made more clear by the findings of this project. 1) Our study demonstrates the ability of proteomics, phospho-proteomics and kinase activity scores to elucidate active signaling processes within tumors. Applying these capabilities to tissue samples in a clinical trial context could yield valuable information regarding the biology of individual tumors that respond to a given therapy. 2) As these determinant proteomic signatures are identified,

MRM signatures can be developed to identify patients in real time whose tumors display particular biologic features and thus may respond to a treatment. 3) Histologically similar tumors are frequently treated differently in pediatric and adult settings. While they often differ in their genomic features, this study has shown that they do not always drive biology. A unique opportunity building off of this work will be to use proteomic platforms to interrogate tumors whose incidence spans a large age range in order to answer questions regarding how biology changes across the spectrum and whether treatments can be realigned for maximum benefit. In summary, this large multi-omics study of pediatric brain tumors represents an entrée for the integration of proteomics into data science modelling of pediatric cancer and as such, it sets the stage for more applied research to come.

## STAR METHOD

### EXPERIMENTAL MODEL AND SUBJECT DETAILS

**Patient selection for the discovery cohort**—The samples were obtained from the Children's Brain Tumor Tissue Consortium (CBTTC) at the Children's Hospital of Philadelphia (CHOP). The patient selection was built based on specimen availability and defined two broad classes of tumors: (1) High grade tumors driven by epigenetic dysregulation (HGG, DIPG, ATRT and/or other embryonal tumors) and (2) Low grade tumors defined by receptor-tyrosine kinase and MAPK signaling alterations including kinase fusions. Additional associated clinical determinants for cohort selection included: (A) Tumor histologies for which there is more than one therapeutic standard of care and for which a multidimensional proteogenomic analysis could further inform an assessment of therapeutic response; (B) Tumor histologies for which genomic alterations and/or classification have failed to provide differential prognosis; (C) Tumor cohorts for which comprehensive profiling could inform the course of metastasis. These considerations led to the selection of 226 samples from 204 pediatric subjects treated surgically and clinically at the Children's Hospital of Philadelphia for whom deep longitudinal, clinical data is available.

**Sample collection for the discovery cohort**—Samples were collected at the time of surgery (217 samples) or autopsy (9 samples), flash-frozen, and stored in BioRC (Biorepository Resource Center) at Children's Hospital of Philadelphia. Frozen tissue pieces ~75mg were cut off using disposable scalpels on dry ice and delivered to Fred Hutchinson Cancer Research Center for sample preparation for proteomics profiling. ~20 mg frozen tissue and up to 0.4–1ml of blood was used for nucleic acid extractions, which were performed at the Biorepository Resource Center at Children's Hospital of Philadelphia.

**Sample collection for the HGG validation study**—For the validation studies, the specimens from 41 patient subjects 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. Among the 41, 18 were part of the discovery cohort who had remaining tumor materials. 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. ~30mg tissue pieces were cut/chipped off using disposable scalpels on dry ice and delivered to Fred Hutchinson Cancer Research Center for sample preparation for proteomics profiling.

## METHOD DETAILS

**Nucleic acid extractions, WGS and RNAseq**—Tissues were lysed with Qiagen TissueLyser II (Qiagen) using 5 mm steel beads (cat# 69989, Qiagen) 2×30 sec at 18Hz settings, and processed with CHCl3 extraction and run on the QiaCube automated platform (Qiagen) using the AllPrep DNA/RNA/miRNA Universal kit (cat# 80224, Qiagen). Thawed blood was RNase A (cat#, 19101, Qiagen) treated and processed using the Qiagen QIAsymphony automated platform (Qiagen) using the QIAsymphony DSP DNA Midi Kit (cat# 937255, Qiagen). DNA and RNA quantity and quality was assessed by PerkinElmer DropletQuant UV-VIS spectrophotometer (PerkinElmer) and an Agilent 4200 TapeStation (Agilent, USA) for RINe and DINe (RNA Integrity Number equivalent and DNA Integrity Number equivalent respectively). Library preparation and sequencing was performed by the NantHealth sequencing center. Briefly, DNA sequencing libraries were prepared for both tumor tissue and matched-germline (blood) DNA using the KAPA Hyper prep kit (cat# KK8541, Roche); Whole genome sequencing (WGS) was performed at an average coverage of 60X for tumor samples and 30X for matched-germline. The panel tumor sample was sequenced to 470X and the normal panel sample was sequenced to 308X. Tumor RNA-Seq libraries were prepared using KAPA Stranded RNA-Seq with RiboErase kit (cat# KK8484, Roche). RNA samples were sequenced to an average of 200M reads. All sequencing was performed on the Illumina HiSeq platform (X/400) (Illumina) with $2 \times 150$bp read length.

**Somatic Mutation and CNV calling**—Strelka2 (Kim et al., 2018b) v2.9.3 was run for canonical chromosomes (chr1–22, X,Y,M) using default parameters and the resulting VCF was filtered for PASS variants. Gene level mutation status were summarized based on somatic mutations detected in coding regions, having minimum sequencing depth of 30, and minimum alternative variant count of 5.

CNVkit v. 2.9.3 (Talevich et al., 2016) was run in batch wgs mode, paired tumor-normal, using the hg38 annotation reference from UCSC (http://hgdownload.soe.ucsc.edu/goldenPath/hg38/database/refFlat.txt.gz). All output files, such as seg, gain/loss, and scatter/diagram plots were generated using CNVkit's export and other built-in functions.

**RNAseq data preprocessing**—STAR v2.6.1d (Dobin et al., 2013) was used to align paired-end RNA-seq reads against the ENSEMBL GENCODE 27 "Comprehensive gene annotation" reference (https://www.gencodegenes.org/human/release_27.html). RSEM v1.3.1 (Li and Dewey, 2011) was used to generate both FPKM and TPM transcript- and gene-level expression values. Then, log2(x+1) transform was applied and samples with replicates were averaged.

### Proteomic experiments for Discovery Cohort

**11-Plex Preparation:** Sample preparation for MS analysis was performed as described previously (Navarrete-Perea et al., 2018). Lysates were prepared from 226 Cryo-pulverized human Pediatric Brain Tumor samples in lysis buffer (6M Urea, 25 mM Tris, pH8.0, 1 mM EDTA, 1 mM EGTA, Sigma Protease Inhibitor Cat# P8340, Sigma Phosphatase Cocktail Cat# P5726, Sigma Phosphatase Cocktail Cat# P0044) supplied by Fred Hutch. Lysates were reduced with 5 mM neutralized TCEP (Pierce, #77720) for 15 min., alkylated with 10 mM Iodoacetamide (Sigma, #A3221) for 30 minutes in the dark and quenched with 5 mM Dithiothreitol (Thermo Scientific, #20291) for 15 mins. Protein was precipitated with methanol-chloroform, and the protein pellet was resuspended in 200 mM EPPS (pH 8.0). The samples were digested sequentially with Lys-C protease (Wako, 129–02541, 2 mg/mL Stock) at a 100:1 protein-to-protease ratio with constant shaking overnight at room temperature followed by Trypsin (Pierce, 90305, 1 mg/mL stock) digestion at a 100:1 protein-to-protease ratio for another 6 h at 37°C. Digested peptides were assayed for peptide concentration with Pierce™ Quantitative Colorimetric Peptide Assay (#23275) as per Manufacturer's protocol.

**11-plex Experimental Layout:** Proteome and Phosphoproteome analysis of pediatric brain cancer samples were structured as TMT11-plex experiments. 226 unique samples plus a few replicates and QC samples were arranged in twenty-three 11-plex experiments with 10 individual samples occupying the first 10 channels of each experiment and the 11th channel being "Bridge Channel" i.e. Common Reference Sample, used for quantitative comparison across all sample sets. To prepare the bridge channel which broadly represents the population of pediatric brain cancer samples in our experiment, digested peptides from indicated samples were pooled together.

**TMT Labeling of Peptides and Quality Check:** About 100 μg of digested peptides per sample were labeled with TMT11-plex reagent according to the manufacturer's instructions (Thermo Scientific, Pierce Biotechnology, Germany). About 2 μg of each sample from each 11-plex experiment was removed and combined in 100 μl of 1% formic acid (FA) for a quality control check. The remaining samples were frozen immediately at −80°C for future quenching and HPLC fractionation. The combined samples in 100 μl of 1% FA from each 11-plex experiment were desalted by StageTip containing 4 small (0.9 mm) discs of 3M™ Empore™ C18 material following standard procedure. Eluted material was dried by speedvac, resuspended in 5% ACN/5% FA and analyzed by a mass spectrometer to check (a) digestion efficiency, (b) labeling efficiency & (c) summed signal-to-noise ratios among samples. As a standard of quality check (QC), minimum of 97% labeled MS/MS spectra, a maximum of 5% missed cleavage rate, and summed signal-to-noise ratio variations of <1.5-fold within each plex were required to proceed further. Following successful QC checks, unwanted TMT labeling of tyrosine residues was reversed with a final concentration of ~0.3% (v/v) hydroxylamine (Sigma, 467804, 50% stock) for 15 mins and finally quenched with 50% TFA to a final concentration of ~0.5% (v/v). Labeled peptides from each of the twenty three 11-plex experiments were combined into 23 samples, acidified, and subsequently desalted on C18 Sep-Pak columns. Eluates were dried by SpeedVac in preparation for Phosphopeptide Enrichment.

**Phosphopeptide Enrichment:** High-Select™ Fe-NTA Phosphopeptide Enrichment Kits (Thermo Scientific™; #A32992) were used for phosphopeptide enrichment step ("mini-phos") as per protocol described (Navarrete-Perea et al., 2018). Lyophilized labelled peptide sample was completely dissolved in 200 μL of Binding/Wash buffer with vortexing, and pH was confirmed to be below 3. Samples were loaded onto spin columns equilibrated per Manufacturer`s method and mixed by gentle tapping until the resin was in suspension. Samples were incubated for 30 minutes at room temperature with gentle mixing every 5 minutes. Following incubation, the spin columns were placed in a microfuge tube, centrifuged at 1000×g for 30 seconds, and washed thrice with a Binding/Wash buffer. All flow-through fractions were collected in the same tube, desalted, dried, resuspended and directed to basic-pH HPLC for global proteome analysis. Phosphopeptide-bound spin columns were placed in a new microfuge tube, containing 1% FA and Phosphopeptides were eluted with an Elution buffer and dried immediately by speedvac. For phosphoproteome analysis, phosphopeptide enriched samples were resuspended in 1% FA, desalted by stage tip and eluted into Agilent deactivated glass vial inserts, dried by speedvac and finally phosphopeptide enriched samples were resuspended with 10 μL of 5% FA and made ready to be analyzed by LC-MS/MS analysis.

**Offline fractionation of peptides:** To reduce sample complexity, peptide samples were separated by high pH reversed phase (RP) fractionation. Phosphopeptide Flow-Through fractions were checked for pH (pH < 3) and desalted on C18 Sep-Pak columns. Eluates were dried in a Speed Vacuum Concentrator, reconstituted in 500 μL of Buffer A (10 mM ammonium bicarbonate, 5% Acetonitrile, pH 8), loaded onto an Agilent 300Extend C18 column (3.5 μm bead size, 4.6 mm ID and 220 mm long), and separated on an Agilent 1200 HPLC instrument at a flow rate of 0.6 mL/min with a 60 min linear gradient from 13% to 42% buffer B (10 mM ammonium bicarbonate, 90% ACN, pH 8) into a total of 96 fractions. Each fraction contained ~500uL, at ~37 seconds per fraction. All 96 fractions were consolidated into 12 final fractions by column, desalted with stage-tip, resuspended with 10 μL of 5% ACN - 5% FA and made ready to be analyzed by LC-MS/MS analysis.

**Mass Spectrometry (MS) Instrument:** Global proteome analyses were performed on an Orbitrap Fusion™ Tribrid™ Mass Spectrometer and phosphoproteome analyses were performed on an Orbitrap Fusion™ Lumos™ Tribrid™ Mass Spectrometer (Thermo Scientific™) both in-line with a NanoSpray Flex NG ion source using MS3-based TMT centric mass spectrometer method. Both the instruments are online with a Liquid Chromatography System (EASY-nLC™ 1200 System).

**Online Liquid Chromatography:** Online separation was performed on a nano-flow UHPLC EASY-nLC™ 1200 system (Thermo Scientific™). In this set up, the LC system with a column in (IDxL 20 mm × 550 mm; LC 560) from sample valve and a column out (IDxL 75 mm × 550 mm; LC 562) to waste valve, a Fused Silica Capillary Tubing, that delivers sample to MS were connected via a stainless-steel cross union (Micro-Cross Assay-SS 360μm UH-906/ Western Analytical products). A platinum wire was used to deliver electrospray source voltage. The column was heated to 60°C using a column heater sleeve (Phoenix-ST) to prevent over-pressuring of columns during UHPLC separation. The

capillary tubing (Inner Diameter: 100μm, Outer Diameter: 375μm) was pulled to an opening of 10 mm and packed in-house with 2.6 mm beads (90 A Pore diameter, Thermo Scientific) slurry made in Buffer B (90% Acetonitrile, 0.1% FA). Each analysis, ~1 mg of peptide in a 1–5 ml injection volume based on the sample dilution, and for each Phosphoproteome sample in a 5 ml, were loaded onto the column in Mobile phase, comprised of 0.1% FA (Buffer A). LC-MS/MS method consisted of an initial 10 min column-equilibration procedure and a 20 min sample-loading procedure, both at 800 bar. For global proteome analyses, the peptides were separated using a 150 min gradient of 5 to 42% Acetonitrile in 0.1% FA and respective flow rate was adjusted to separate the fraction within a pressure difference of 300 – 400 bar. Phosphopeptides were separated over 160 min with gradient of 3 to 30 % Acetonitrile in 0.1% FA.

**Mass Spectrometry Analyses:** For data-dependent experiments (MS2 and MS3), all instrument operational parameters were specified through the instrument method editor. Data-dependent acquisition was performed using Xcalibur v2.1 software in positive ion mode at a spray voltage of 2.6 kV and 300oC ion transfer tube temperature. For 276 Peptide fractions analyses, MS1 Spectra was detected with Orbitrap at a resolution of 120K, scan range (m/z) being 350 – 1350 and AGC target being 1.0e6 with 50 ms maximum ion injection time. For MS2 analysis, top ten precursors were selected with peptide as monoisotopic peak determination, intensity threshold of 1.0e3, and charge state screening was enabled to include only precursor charge states 2–6. Peptides that triggered MS/MS scans were dynamically excluded from further MS/MS scans for 90 sec, with a ± 10 ppm mass tolerance. Perform dependent scan on single charge state per precursor only and Exclude within the cycle were enabled. In data-dependent charge specific MS2 analysis, ions were first isolated by Quadruple with an isolation window of 0.7 or 0.5 (based on instruments used) and activated at ion Trap with CID collision Energy being 35 % in 10 ms and activation Q of 0.25. Ion trap detection were set to normal scan range mode with rapid ion trap scan rate, 9.0e3 AGC target and 80 ms ion injection time. Following the acquisition of each MS2 fragment ion, precursors were selected with a mass range (m/z) between 400 – 2000 with a mass exclusion width 50 (low) and 5 (high). About 10 precursor fragment ions were simultaneously isolated by SPS selection at a time for every MS3 precursor population, which then fragmented by HCD with HCD collision energy of 55% and fragmented reporter ions with a normal scan range mode were analyzed in Orbitrap at a resolution of 50K, AGC Target 1.0e5 and 120 ms maximum ion injection time. To further minimize the influence of co-eluting species, peptides with isolation specificities less than 1.2 for + 2 charge, 1.0 for + 3 charge and 0.8 for 4 – 6 charge.

Each of the 23 Phosphopeptide enriched samples were directed to both CID and High Resolution HCD activation using similar multinotch MS3-based TMT methods. For CID activation, both multistage activation and injections for all available parallelization time were enabled with ion trap scan rate set to Turbo and neutral loss mass to 97.9763. For HCD activation during MS2 analysis, HCD collision energy was set to 32%. Both CID and HCD activation are considered as two fractions from each of the 23 Phosphopeptides enriched samples, totaling 46 Peptide fractions.

**Protein Identification and Quantification**—Proteomics processing of whole proteome and phosphopeptide-enriched datasets was performed as described previously (Clark et al., 2019; Djomehri et al., 2020). MSFragger version 20190628 (Kong et al., 2017) was used to search a CPTAC harmonized RefSeq protein sequence database appended with an equal number of decoy sequences. MS/MS spectra were searched using a precursor-ion mass tolerance of 20 ppm, fragment mass tolerance of 0.7 Da, and allowing C12/C13 isotope errors (-1/0/1/2/3 for global, and 0/1/2/3 for phosphopeptide-enriched). Cysteine carbamidomethylation (+57.0215) and lysine TMT labeling (+229.1629) were specified as fixed modifications, and methionine oxidation (+15.9949), N-terminal protein acetylation (+42.0106), and TMT labeling of the peptide N terminus. For the whole proteome database search, TMT labeling on Serine residues was also specified as a variable modification. The search was restricted to fully tryptic peptides, allowing up to two missed cleavage sites. Phosphopeptide-enriched searches also included the phosphorylation modification of serine, threonine, and tyrosine residues (+79.9663). The search results were then processed using the Philosopher toolkit version v1.2.3 (da Veiga Leprevost et al., 2020), including PeptideProphet (Keller et al., 2002), PTMProphet (Shteynberg et al., 2019), and ProteinProphet (Nesvizhskii et al., 2003). The data was filtered to 1% PSM-level (for each 11-plex), and 1% protein-level (global) FDR using Philosopher filter command. TMT-Integrator version v1.0.4 (http://tmt-integrator.nesvilab.org/) was used for generation of quantification matrices as described previously (Clark et al., 2019; Djomehri et al., 2020), except its parameters were adjusted to process 11 channels, a minimum peptide probability of 0.5 for quantification, and minimum site localization probability of 0.75 (phosphopeptide-enriched datasets only). Quantification results (log2 ratios) were summarized at protein and gene levels, and for phosphopeptide enriched data also at the site-level.

**Preprocessing of TMT proteomic data**—8802 unique genes and 18235 phosphosites were identified and quantified from the global proteomic and phopshoproteomic experiments. The Global normalization were performed on gene-level abundance matrix (log2 ratio) for global proteomic and site-level abundance matrix (log2 ratio) for phosphoproteomic data. Specifically, each sample were shifted to have the same median, and scaled to have the same median absolute deviation.

We then applied an 'Intra TMT-multiplex t-test' to detect and remove outlier TMT multiplexes for each protein/phosphosite. For each TMT multiplex, we performed t-test of the protein/phosphosite abundance of samples inside against the protein/phosphosite abundance of samples outside the multiplex. TMT multiplexes with a p-value lower than 10e-7 were flagged as outliers and removed from the dataset. Accordingly, a total of 164 and 156 multiplexes or 1612 and 1557 data points were removed in the global protein abundance and phosphosite datasets respectively.

Before performing any downstream analysis, we applied diagnosis-specific batch correction on both global and phospho abundance to remove the technical difference between different TMT 10-plex. For each data type, batch correction was performed on the subset of markers with more than 50% observed in at least one of the subtypes. After filtering markers with missing rates >50% in all subtypes, there were 6429 genes and 4988 phosphosites with an

overall missing rates of 20.2% and 40.8% for the diagnosis-wide global protein and phospho abundance datasets respectively.

As a first step to batch correction, we performed KNN imputation separately on the data from each diagnosis using the "impute.knn" function from the "impute" R package. After merging the data across diagnoses, we then applied the R tool ComBat, with the tumor diagnosis as covariate to remove batch effects (Johnson et al., 2007). Finally, we replaced the missing data structure from before KNN imputation.

For the formal imputation of missing values, we adopted a novel tool DreamAI (Ma et al., 2020) (https://github.com/WangLab-MSSM/DreamAI), an ensemble algorithm developed during the NCI-CPTAC Dream Proteomics Imputation Challenge (https://www.synapse.org/#!Synapse:syn8228304/wiki/413428). Imputation was done: 1) separately on the data from each tumor type, and 2) across the entire dataset including all tumors. Tumor-subtype specific imputation was done for the subset of markers with missing <50% in each subtype. Subtype-wide imputation was done for the subset of markers that appeared in at least 50% of samples in any one subtype (the same set of markers used in the batch correction). Finally, for the phospho abundance dataset, we filtered out 440 additional markers associated with cold-regulated ischemia genes.

**QC check for proteogenomic profiles**—Integration of these multi-layers of omics data enhances our understanding about complex molecular mechanisms in biological systems. However, unintended errors in annotations and sample sables often occur in generation or management of large-scale data (Alyass et al., 2015). Since integrative analysis based on error-containing data could provide wrong scientific conclusions, data quality and sample-labeling check is a critical QC step before actual integration. In this study, we performed systematic quality control procedure to confirm that all annotations in clinical information and sample names are consistent as annotated.

1. *Diagnosis type check and filtering* Among the 226 samples, 7 samples were identified to have either incorrect or ambiguous histologic diagnosis based on an independent clinical report review and were removed from the downstream analysis.

2. *Gender label check* Expression of two gender representative genes, XIST and RPS4Y1 from chromosomes X and Y, respectively (Staedtler et al., 2013), were used to infer genders based on RNAseq data.

3. *Genotype mapping check based on WGS and RNAseq data* To ensure the highest data quality, genotype mapping analysis was performed to flag samples with potential contamination, low sequencing quality, or sample labelling issues. Specifically, using NGSCheckMate (Lee et al., 2017), the genotype correlation were compared between paired tumor WGS vs normal WGS as well as tumor WGS vs tumor RNAseq profiles for all patients. The tool utilized 20,000+ common SNP sites based on dbSNP138, and a stringent cutoff of 0.8 was applied to flag low-quality or contaminated samples.

**4.** *Proteo-genomic sample labelling mapping* We employed similar procedures applied to our recent kidney cancer study (Clark et al., 2019) to confirm that RNAseq, global proteomics, and phosphoproteomic data with the same labels were from the same individuals (Yoo et al., 2014). Cis pairs among global proteomics, phosphosite proteomic, and RNAseq data were determined based on their correlation strength (cis correlation>0.6) and 832, 1341, and 521 pairs were selected for global-phosphosite, RNAseq-global, and RNAseq-phosphosite alignments, respectively. Then the values of the selected features (genes or proteins) were rank-transformed to evaluate sample-wise similarity scores. If two profiles of a sample are matched, the similarity score between the two profiles is expected to be significantly higher than the score of random pairs. Based on this approach, we identified potentially mis-aligned samples from pairwise alignment among global proteomic, phosphoproteomic, and RNAseq data.

For flagged samples, same subject tumor tissue and blood specimen DNA was further extracted and sent to Guardian Forensic Sciences (Abington, PA) for short tandem repeat (STR) testing using the GenePrint 24 assay (Promega, #B1870). Pattern of amplified polymorphic loci was used for matching analysis between tissue and blood for each case.

After filtering data files according to all above quality assessments, the resulting data sets consisting of 218 global proteomics profiles, 217 phosphoproteomics profiles, 188 RNAseq profiles, 200 mutation profiles and 190 CNV profiles were considered for downstream analyses (Table S1).

## Proteomics experiment of the validation cohort.

**Protein Extraction and Lys-C/Trypsin Tandem Digestion:** Approximately 50 mg of each of brain tumor tissues were cryopulverized and lysed separately in 800 μL 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 g for 10 min at 4 °C and protein concentrations were determined by BCA assay and adjusted to approximately 1.5 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 to 2 M urea concentration with 25 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, aliquot of sequencing grade modified trypsin (Promega, V5117) at 1:25 enzyme-to-substrate ratio was added to the samples and further incubated at 25 °C for 14 h. The digested samples were then acidified with 100% formic acid to 1% of final concentration of formic acid and centrifuged for 15 min at 1,500 g to clear digest from precipitation. Tryptic peptides were desalted on C18 SPE (Waters tC18 SepPak, WAT054925) and dried using Speed-Vac.

**TMT-11 Labeling of Peptides:** Desalted peptides from each sample were labeled with 11-plex TMT reagents. 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 100 adult GBM tumor and 10 GTEx normal samples (Wang et al., 2020)) 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 desalted on tC18 SepPak SPE columns.

**Peptide Fractionation by 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 × 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:** $Fe^{3+}$-NTA-agarose beads were freshly prepared using the Ni-NTA Superflow agarose beads (QIAGEN, #30410) for phosphopeptide enrichment. For each of the 12 fractions, peptides were reconstituted to 0.5 μg/μL IMAC binding/wash buffer (80% acetonitrile, 0.1% trifluoroacetic acid) and incubated with 10 μL of the 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.

**LC-MS/MS Analysis:** Fractionated samples prepared for whole proteome and phosphoproteome 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. × 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. 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 whole 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. × 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 $4 \times 105$) 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 $1 \times 105$) 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.

**Quantification of TMT Whole Proteomic Data:** 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 et al., 2008; Kim and Pevzner, 2014) to match against the RefSeq human protein sequence database downloaded on June 29, 2018 (hg38; 41,734 proteins), combined with 264 contaminants (e.g., trypsin, keratin). The partially tryptic search used a ± 10 ppm parent ion tolerance, allowed for isotopic error in precursor ion selection, and searched a decoy 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. (Monroe et al., 2008). Next, PSMs passing the confidence thresholds described above were linked to the extracted reporter ion intensities by scan number. The reporter ion intensities from different scans and different bRPLC fractions corresponding to the same gene were grouped. Relative protein abundance was calculated as the ratio of sample abundance to reference abundance using the summed reporter ion intensities from peptides that could be uniquely mapped to a gene. The pooled reference sample was labeled with TMT 126 reagent, allowing comparison of relative protein abundances across different TMT-11 plexes. The relative abundances were log2 transformed and zero-centered for each gene to obtain final relative abundance values. Small differences in laboratory conditions and sample handling can result in systematic, sample-specific bias in the quantification of

protein levels. In order to mitigate these effects, we computed the median, log2 relative protein abundance for each sample and re-centered to achieve a common median of 0.

**Quantification of Phosphopeptides:** Phosphopeptide identification for the phosphoproteomic data files were performed as in the whole 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. To account for sample-specific biases in the phosphoproteome analysis, we applied the correction factors derived from median-centering the whole proteomic dataset. Preprocessing of the proteomic tables of the Project Hope sample analysis were performed in the same fashion in the pediatric sample analysis described above.

**Targeted Mass Spectrometry Methods**—For targeted mass spectrometry measurements, tissue lysates were reduced, alkylated with iodoacetamide, and digested by the addition of trypsin at a 1:50 trypsin:protein ratio (by mass), as previously described (Whiteaker et al., 2018). After 2 hours, a second trypsin aliquot was added at a 1:100 trypsin:protein ratio and incubated overnight at 37°C with shaking. After 16 hours, the reaction was quenched with formic acid (final concentration 1% by volume). A mix of stable isotope-labeled peptide standards was added to the digest at 80 fmol/mg per peptide.

Peptide immunoaffinity enrichment was performed as previously described (Zhao et al., 2011), using a mixture of 50 antibodies crosslinked on protein G beads targeting 75 peptides (21 modifications, 40 proteins). LC-MRM was performed as previously described (Whiteaker et al., 2018).

**Targeted MRM Assay Characterization:** Response curves were generated in a background matrix of pooled brain tumor lysates. Five hundred microgram aliquots of the pooled lysate were digested by trypsin, and the heavy stable isotope-labeled peptides were added to aliquots in triplicate by serial dilution covering the amounts 1000, 200, 40, 8, 3.2, 1.28, 0.512, 0.205 fmol/mg with light spiked into the pool at 80 fmol/mg. Blanks were prepared using a background matrix with light peptide (no heavy spike). All points were analyzed in triplicate (including peptide addition, immunoaffinity enrichment, and mass spectrometry). Data analysis was performed using Skyline. The Lower Limit of Quantification (LLOQ) was obtained by empirically finding the lowest point on the curve with a CV <20% in the curve replicates. All measurements were filtered by the LLOQ (i.e. all measurements were required to be above the LLOQ). The upper limit of quantification (ULOQ) was determined by the highest concentration point of the response curve that was maintained in the linear range. For curves that maintained linearity at the highest concentration measured, the ULOQ is a minimum estimate.

Repeatability was determined using the same pooled lysate matrix used to generate the response curves with heavy peptides spiked in at three concentrations (0.8, 80, 800 fmol/mg) and light peptides added at 200 fmol/mg. Complete process triplicates (including digestion,

capture, and mass spectrometry) were prepared and analyzed on five independent days. Intra-assay variation was calculated as the mean CV obtained within each day. Inter-assay variation was the CV calculated from the mean values of the five days.

**Targeted MRM Data Results:** The median LLOQ was 1.6 fmol/mg and the median linear range >2.8 orders of magnitude. In repeatability experiments, the median CV at the medium spike level was 8.6% (intra-assay CV) and 26% (inter-assay CV).

Each data point is the peak area ratio (light:heavy) filtered by the LLOQ. The unfiltered data points are also available in the table. For each sample 500 ug aliquots were analyzed in complete process replicate (including digestion, capture, and mass spectrometry). The number of replicates available for processing was determined by the amount of lysate available. Overall, 68 out of the 75 peptide analytes were detected in >50% of the samples above the LLOQ. Five peptides were not detected in any samples (TNF10.pan.NGELVIHEK, ATM pS2996, ATMpS367, RIF1 pS1542, RIF1.pan.ASQGLLSSIENSESDSSEAK). For peptides with replicates available, the median CV was 12.9%. The correlation of peak area ratios for peptides originating from the same protein was high (RPTOR: R2=0.9047, ERBB2: R2=0.9536, K25: R2=0.9964) indicating good quality for multiple measurements of the same protein.

## QUANTIFICATION AND STATISTICAL ANALYSIS

**Kinase Activity Score Calculation—**Substrates of every kinase were collected from the PhosphoSitePlus database (version 052819). We only considered kinases with at least five substrates observed in our phosphoproteomic data. To calculate the kinase activity score for each sample, we run a Wilcoxon rank sum test comparing the abundance of substrates of a particular kinase with that of the remaining phosphosites observed in our data. This test was performed for each kinase and each of the 209 samples (i.e., excluding the post-mortem samples). The normalized test statistic of the Wilcoxon test was utilized as the activity score for each kinase.

### Consensus clustering analysis

**Proteomic cluster:** Consensus clustering was performed to identify proteo-typical clusters of childhood brain tumors. Based on gene level global proteomics data, features (genes) were first filtered according to the coefficient of variation (CV) and standard deviation (SD) across samples. Specifically, the CV was calculated using the raw intensity data; features with CV less than 0.1 were filtered resulting in the exclusion of 583 genes for consideration; finally, 3000 genes with the highest standard deviation across 218 samples were selected for clustering.

Consensus clustering was performed using the ConsensusClusterPlus package in R (Wilkerson and Hayes, 2010). Prior to clustering the data matrix was scaled so that each peptide had a mean 0 and a sd of 1 across samples. K-means clustering based on an Euclidean distance metric was conducted across 500 repetitions for cluster numbers ranging from 2 through 10 using otherwise default parameters.

**Phosphoproteomic and transcriptomic cluster:** To compare proteomic clusters with those derived from alternative -omic data types, RNA-seq and phosphosite clusters were identified using a similar procedure. Specifically, phosphosite data from 217 samples were clustered using the 3000 phosphosites with the largest SD after first filtering 830 phospho-sites with CV>0.1. For the clustering of RNA-seq data, ½ of the genes with the highest standard deviation were selected corresponding to 9104 features from 188 tumor samples.

Inspection of the CDF distribution, as well as patterns of concordance across data types and with histological diagnosis, led to the selection of 8 clusters for further analysis.

**Comparison across single-omic clusters:** To evaluate and compare the cohesiveness of allocations derived using single-omic clustering, we utilized silhouette scores. Silhouette scores, which measure the similarity of a given sample to the other samples in the same cluster, were calculated using the silhouette function from the 'cluster' package in R. For each single-omic dataset, the Euclidean distance matrix used in the consensus clustering and the respective single-omic clustering allocations were included as inputs to the silhouette function.To further compare clustering allocations across single-omic datasets and with histological diagnoses, the percentage of each diagnostic type falling into each cluster was calculated for each set of allocations. This data is summarized in Figure S1B.

**Survival analysis for proteomic clusters**—The association of proteomic clusters and overall survival was evaluated using a Cox model based on 198 patients with surgical samples and overall survival information (Figure 1C). Overall survival values were truncated to a maximum of 3750 days.

**Stemness Score**—The stemness indices are used for assessing the degree of oncogenic dedifferentiation, as previously described (Malta et al., 2018). In other words, stemness can be considered (Malta et al., 2018). In other words, stemness can be considered to be the ability of the tumor to phenocopy a normal stem cell. Higher values for stemness indices were associated with biological processes active in cancer stem cells and with greater tumor dedifferentiation, as reflected in histopathological grade (Malta et al., 2018). Recently several signaling pathways associated with stemness have been reported for each of the mentioned CBTTC PBT diagnoses (Chang et al., 2017; Liu et al., 2019; Meel et al., 2018).

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 PCBC dataset. To calculate mRNA based stemness index (mRNASi), we used the FPKM (Fragments Per Kilobase Million) mRNA expression values for all the 188 CBTTC PB tumors. We used the function TCGAanalyze_Stemness from the package TCGAbiolinks (Colaprico et al., 2016) and following our previously-described workflow (Mounir et al., 2019), with "stemSig" argument set to PCBC_stemSig.

**Proliferative Index—**Proliferative index was calculated based on gene expression data of 40 genes contained in the proliferation gene signature from Yuan et al, (Yuan et al., 2018). The proliferative index was computed via ssGSEA score using the package GSVA (Hanzelmann et al., 2013).

**Investigation of two subtypes of Cranio—**To better characterize the biological features differentiating the CP allocated to different proteomic clusters and to investigate the hypothesis that CP allocated to the *Cranio/LGG BRAF$^{V600E}$ (C4)* may respond to MEK inhibitor treatment, regression analyses were performed using the global proteomic data to identify markers differentially expressed between C4 and C8 with CTNNB1 status accounted for as a covariate. Gene set enrichment tests were performed for a MEK inhibition response signature based on 15 genes overlapping between our global proteomic data and a 52 member geneset previously reported to be perturbed by MEK inhibitor treatment in multiple cancer cell lines with *BRAF$^{V600E}$* (Pratilas et al., 2009). The MEK inhibition response gene set was found to be significantly enriched of proteins upregulated in C4 (pvalue=0.05), as is illustrated in the volcano plot in Figure S1C.

**Proteomic Cluster Signature—**To identify proteins and phospho-site markers associated with proteomic clusters, a multiple regression was performed using protein/phospho-site abundances as responses, and binary indicators representing the 8 proteomic-clusters as regressors, with age of specimen diagnosis, gender, as well as treatment and clinical status at sample collection included as covariates. Model fitting was performed without an intercept so that the resulting betas are interpretable as a mean shift relative to all tumors. Results from cluster-specific association testing performed on 6429 protein (N=218), 4548 phosphosite (N=218), and 18209 gene expression (N=188) features are reported in Table S1.

**Pathways analysis for proteomic clusters—**To better characterize proteomic cluster, we sought to identify the biological pathways distinctly associated with each proteomic cluster. First, genes were cluster based on a Z-score matrix (8 columns) summarizing the cluster-specific regression analysis results based on proteomics data. Specifically, each row of the Z-score matrix represents the estimated mean shift of a given protein's abundance in each of the 8 proteomic clusters. Considering between 10–20 clusters, K-means clustering was performed to group genes (N=6353) whose protein abundances were significantly associated (FDR<0.05) with at least one proteomic cluster. Pathway enrichment was performed to test for overrepresentation of biological pathway/gene-set members in each gene group using a one-tailed Fisher's Exact Test. The final number of gene groups (k=14) was chosen to maximize the number of significant pathway associations based on the Hallmark gene sets from MSigDB (Liberzon et al., 2015; Liberzon et al., 2011) Downloaded from http://software.broadinstitute.org/gsea/msigdb/collections.jsp on 02/14/2019. Based on the 14 groups of gene so selected, a comprehensive pathway analysis was further performed using GO (Ashburner et al., 2000), Biocarta, KEGG (Kanehisa et al., 2017), Hallmark (Liberzon et al., 2015), and Reactome (Fabregat et al., 2018) gene set collections.

**Pathway Consolidation via Sumer:** Gene set enrichment results can be difficult to interpret due to significant redundancy of gene membership across collections of gene sets. To aid in the interpretation and reduce redundancy of pathway results, we utilized the Sumer tool (Savage et al., 2019). This tool uses an affinity propagation algorithm to cluster similar pathway gene sets into largely distinct modules. Sumer was run using -log10 p-values from the Fisher test of gene cluster enrichment as weights. Consolidated pathway modules for each gene cluster were identified, based on the top 50 pathways by weight (Table S1); a subset of these along with other pathways with biological relevance in cancer were selected for shown in Figure 1D.

**Immune subtype identification**—The abundance of 64 different cell types were computed via xCell based on transcriptomic profiles (Aran et al., 2017). Therefore, for this analysis, 182 pediatric brain tumor samples with mRNA data were utilized excluding postmortem samples. Table S2 contains the final score computed by xCell of different cell types. Consensus clustering was performed based on only cells which were detected in at least 5% of the patients (adjusted p-value < 1%). This filtering resulted in 35 cell types. A Microglia signature was derived as ssGSEA score (Hanzelmann et al., 2013) based on the following microglia-specific markers: *P2RY12, TMEM119, SLC2A5, TGFBR1, GPR34, SALL1, GAS6, MERTK, C1QA, PROS1, CD68, ADGRE1, AIF1, CX3CR1, TREM2 and ITGAM* (Butovsky et al., 2014; Crotti and Ransohoff, 2016; Haage et al., 2019; Solga et al., 2015). Based on these 36 signatures, consensus clustering was performed in order to identify groups of samples with similar immune/stromal characteristics. Consensus clustering was performed using the R packages ConsensusClusterPlus (Wilkerson and Hayes, 2010) based on z-score normalized signatures. Specifically, 80% of the original pediatric brain tumor samples were randomly subsampled without replacement and partitioned into 5 major clusters using the Partitioning Around Medoids (PAM) algorithm, which was repeated 200 times (Wilkerson and Hayes, 2010) (Fig. 2A, Table S2).

**single cell RNAseq deconvolution Analysis**—We have applied the tool Music (Wang et al., 2019) trained on single cell sequencing data from (Darmanis et al., 2017) to all the mRNA expression values for all 182 tumors considered for the immune subtype analysis. We used the function TCGAanalyze_scRNA (tool = Music, data = GSE84465) from the package TCGAbiolinks (Colaprico et al., 2016) to query, download and prepare the data from (Darmanis et al., 2017) and subsequently obtain microglia, neuronal and oligodendrocytes cell type composition in these tumors. scRNA data were normalized following previously-described workflow (Lun et al., 2016).

**Tumor Purity, Stromal and Immune Scores**—Besides xCell, we utilized ESTIMATE (Yoshihara et al., 2013) to infer immune and stromal scores based on gene expression data (Table S2). To infer tumor purity, TSNet was utilized (Petralia et al., 2018) (Table S2).

**Differentially Expressed Genes and Pathway**—Genes upregulated in each of the five immune clusters were identified based on gene expression data, global proteomic and phosphoproteomic data. For this analysis, imputed proteomic and phosphoproteomic data were utilized. For each data type, every feature vector was normalized by subtracting the

mean and dividing by the standard deviation across 182 samples. Then, for each data type, the expression level of gene/protein/phosphosite $j$ was modeled via

$$x_{i,j} = \sum_{k=1}^{5} \Box \beta_{k,j} \Box 1(i \in I_k) + \epsilon_{i,j} \qquad (1)$$

with $\epsilon_{i,j} \sim N(0, \sigma_j)$, $I_k$ being the set of samples belonging to the $k$-th immune cluster, 1 ($A$) being an indicator function equal to 1 if the event $A$ occurs and 0 otherwise, $\beta_{k,j}$ being the coefficient capturing the association between gene $j$ and the $k$-th immune group. Benjamini adjusted p-values (Benjamini and Hochberg, 1995) can be found in Table S2. For each immune cluster, considering the set of genes up-regulated with Benjamini's adjusted p-value lower than 1%, a fisher exact test was implemented to derive enriched pathways. For this analysis, pathways from the Reactome (Fabregat et al., 2018), KEGG (Kanehisa et al., 2017), Hallmark (Liberzon et al., 2015) and GO (Ashburner et al., 2000) databases were considered and as background the full list of gene/proteins observed under each data type was utilized. For phosphorylation data, a gene was considered upregulated if at least one substrate of the gene was found upregulated based on phosphorylation data at 1% FDR. The pathway analysis results for different data types are contained in Table S2B. Figure 2B contains key pathways significant at 10% FDR for different data types. Given their similarity in terms of enriched pathways, the two cold immune clusters (i.e., Cold-medullo and Cold-mixed) were combined into one category and pathways upregulated in both clusters at 10% FDR were reported in Figure 2B. Pathway scores for 182 pediatric brain tumor samples were computed based on ssGSEA using the R package GSVA and included in Figure 2A (Hanzelmann et al., 2013).

**Microglia and Macrophage Polarization in LGG**—Microglia polarization signatures were constructed with ssGSEA (Hanzelmann et al., 2013) using RNAseq measurements based on genes described in recent literature (Dello Russo et al., 2017; Fumagalli et al., 2018; Krasemann et al., 2017). Specifically, the following gene sets were considered: Proinflammatory (M1) = (*IL1B, TLR4, TNF, NOS2, APOE, CLEC7A, LGALS3, GPNMB, ITGAX, SPP1, CCL2, FABP5, CYBB*); Anti-inflammatory (M2) = (*COQ7, IL4, IL13, IL10, ARG1, TGFB1, SMAD3, HEXB, P2RY12, MERTK, ENTPD1, TMEM119, TGFBR1, CD163, CD206*). M2–0.65*M1 difference was used for Fig. 2G.

**Immune association with BRAF Status in LGG**—For this analysis, we consider the 35 immune/stromal signature from xCell and the microglia signature utilized to perform the consensus clustering, microglia M1 and M2 signatures and antigen presenting machinery Class I and Class II signature. Antigen presenting machinery signature Class I was derived via ssGSEA score (Hanzelmann et al., 2013) using gene expression measurements of *HLA-A*, *HLA-B* and *HLA-C* genes; while class II signature based on the gene expression of *HLA-DPA1*, *HLA-DPB1*, *HLA-DRA*, *HLA-DRB1*, *HLA-DRB5* and *HLA-DQB1*. Each signature was normalized to z-score and then was modeled as function of *BRAF* status (i.e., BRAF wild-type, BRAF fusion and BRAF v600E) via linear model. Table S2 reports p-values for association passing 10% FDR.

**iProFun Based Cis Association Analysis—**We investigated the functional molecular quantitative traits (mRNA, protein, and phosphoprotein abundances) perturbed by CNV, using an integrative analysis tool iProFun (Song et al., 2019). iProFun jointly models the multi-omics outcomes, and enjoys largely enhanced power for detecting significant cis-associations shared across different omics data types; and it also achieved better accuracy in inferring cis-associations unique to certain type(s) of molecular trait(s). Specifically, we considered three functional molecular quantitative traits (mRNA expression levels, global protein abundances, and phosphopeptide abundances) for their associations with CNV measured by log ratios. After removing post-mortem samples, we collected 168 pediatric brain tumor samples with all four platforms measured, and performed iProFun on these samples. Samples from different biopsies of the same subject (e.g. from initial tumor and progressed tumor) were both considered in the analysis. The mRNA expression levels were available for 18,209 genes, the global protein abundance measurements were available for 6,429 genes, the phosphopeptide abundance was available for 4,518 peptides from 1,958 genes, and the CNVs were obtained for 19,374 genes, respectively. All data types were preprocessed to eliminate potential issues for analysis such as batch effects, missing data and major unmeasured confounding effects. For this analysis, imputed proteomic and phosphoproteomic tables were utilized. The mRNA expression levels, global protein and phosphoprotein abundances were also normalized to standard normal distribution. To account for potential confounding factors, we considered age, gender, tumor purity, tumor diagnosis, treatment status at collection and somatic mutation. Tumor purity was determined using TSNet from RNA-seq data as described above.

The iProFun procedure was first applied to a total of 1622 genes measured across all 4 data types (mRNA, global protein, phosphoprotein, CNV). Specifically, we started with traditional linear regression for each of the three outcomes separately: mRNA ~CNV + covariates, global ~ CNV + covariates, and phospho ~ CNV + covariates. Then, the association summary statistics from regressions was taken as input for iProFun to call posterior probabilities of belonging to each of the eight possible configurations ("None", "mRNA only" "global only", "phospho only" "mRNA & global", "mRNA & phospho", "global & phospho" and "all three") and to determine significance associations.

Table S3 presents the significant genes that pass the following 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 QTs. 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 CNV being associated with mRNA expression levels was obtained by summing up the posterior probabilities in the following four association patterns – "mRNA only", "mRNA & global", "mRNA & phospho" and "all three", all of which were consistent with CNV being associated with mRNA expression. Lastly, we calculated empirical FDR via 100 permutations per molecular QTs by shuffling the label of the molecular QTs, and requested

empirical FDR (eFDR) <10% by selecting a minimal cutoff value of alpha that 75%<alpha<100%. The eFDR is calculated by:

$$eFDR = ( \text{Averaged No. of genes with posterior probabilities} > \text{alpha in permuted data} ) / ( \text{Averaged No. of genes with posterior probabilities} > \text{alpha in original data} ).$$

In total, we identified 515 genes whose CNV showed cascading cis-regulation of their mRNA expression levels, global protein and phosphopeptide abundances.

Similarly, iProFun was applied to a total of 6183 genes measured across all 3 data types (mRNA, global protein, CNV) for their cis regulatory patterns in tumors, and 1541 genes whose CNV showed cascading cis-regulation of their mRNA expression levels and global protein abundances. To further visualize the cascading genes from iProFun analysis, we selected a subset of cascading genes which have adequate copy number activity in any of the diagnosis subtypes, and marginally differentiated in protein/phosphosite abundance across different copy number status in the same subtype. We define copy number activity by comparing CNVs with the standard deviation across all samples on the same location: cnv over 1-fold SD was regarded as gain and below negative 1-fold SD regarded as loss. Adequate copy number activity was defined with the total proportion of gain and loss over 25% and either category including at least 2 samples. After categorizing CNV with 3 groups: gain/normal/loss, we tested if protein/phospho abundance was differentially distributed with contrast on gain-to-normal or loss-to-normal by two sample Wilcoxon-test. Genes with p-value below 0.1 in the test under one of the contrasts were indicated as marginal associated with CNV. All of the selected cascade genes were labeled along the genome in Figure S3D and those genes also being reported as druggable targets or oncogenes were listed with their symbols on the same plot.

**Cis-regulation of Somatic Mutations**—We considered genes whose mutation rate is prevalent (at least 6 mutations across 200 tumors) to investigate their associations with their cis mRNA, global and phosphoprotein abundances. A total of 46 genes were therefore considered in this association analysis. For each mutation, we considered the existence of any types of mutation (Yes/No) as primary predictor, mRNA/protein/phosphosite abundance as outcome, and CNV, age, gender, tumor purity and tumor diagnosis types as covariates, and performed linear regressions for their associations.

**Trans association analysis**—For each tumor diagnosis, we investigated the trans associations of its abundant genomic events on all measured mRNA, protein, phosphosite levels that pass QC procedures. For each subtype we calculated the chromosome arm-level cnv activity level using similar criterion as described in iProfun analysis. We compared arm-level CNVs in a specific subtype with their standard deviation across all samples on the same arm, calculated total proportion of gain (over 1-fold SD) and loss (below negative 1-fold SD). The proportion over 25% represented adequate activity in the chromosome arm region. We selected 5 diagnosis subtypes which contains at least 1 active CNV region to test trans associations. We also consider the trans association between two signature mutations and their highly enriched subtypes. Specifically, we considered association on chromosome arm *6q*, *17p*, *17q*, and *22q* in atypical teratoid rhabdoid tumor; *CTNNB1* mutation and *11p*

in CP; *1q* and *8q* in EP; *7p* in ganglioglioma; *NF1* mutation, *1q, 6q, 7p, 9p, 9q, 11p, 13q, 14q, 16q, 17q* and *21p* in high grade astrocytoma; *1q, 7p, 7q, 8q, 10q, 11p, 11q, 16q, 17p, 17q* and *18p* in MB. For each of these genomic events, we investigated their association with all mRNA levels, protein abundances and phosphosite abundances among patients with the corresponding diagnosis, using unadjusted linear regression. Additional covariates were not considered due to small sample sizes in subtypes. We reported significant trans associations if FDR < 0.1.

To further understand the biological impact of the trans-regulations, we tested enrichment of positively regulated or negatively regulated gene set in pathways with fisher exact test. Enrichment test was performed on both RNA trans-regulated genes and protein trans-regulated genes. In this test, pathways from the Reactome (Fabregat et al., 2018), KEGG (Kanehisa et al., 2017) and GO (Ashburner et al., 2000) databases were considered and as background the full list of gene/proteins observed under each data type was utilized. Some pathways were enriched by trans-regulated genes in protein but not in RNA. For example, members of the "Cell Cell Contact Zone" pathway (purple) are enriched in the set of proteins upregulated in *CTNNB1* mutant samples; while "Coagulation" pathway is enriched in proteins downregulated in *CTNNB1* mutant samples.

**Kinase Activity across different histologies—**Kinase activity scores were calculated following the strategy illustrated in section "Kinase Activity Score Calculation". Table S4 contains the kinase activity for all diagnosis. For this analysis, we used proteomics and phosphoproteomic imputed data. The activity of each kinase was modeled as a function of the diagnosis indicator and the treatment information via a linear regression. Given the impact of post-mortem collection on proteogenomic data, post-mortem samples have been excluded from the analysis. P-values were adjusted for multiple comparison via Benjamini & Hochberg adjustment (Table S4). In addition, for each diagnosis, the correlation between kinase activity and global abundance is reported in Table S4.

**Kinase-Substrates Association—**To discover the phosphorylation events that were relevant to pediatric brain tumors, we utilized the phosphosite-level data to examine the overall relationship between kinase global abundance and phospho-abundance with targeted sites. Given the impact of post-mortem collection on proteogenomic data, post-mortem samples have been excluded from the analysis. For this analysis, we used proteomic and phosphoproteomic imputed data. For each diagnosis, only kinases and phosphosites measured for at least 50% of the samples have been considered in the analysis. Since ATRT and MB tumors were merged into one group of samples, only kinases and phosphosites observed in at least one diagnosis (i.e., ATRT and MB) for more than 50% of the samples were utilized. For this analysis, experimentally validated kinase-substrate associations were considered from PhosphoSitePlus (Table S4) (Hornbeck et al., 2015). This filtering resulted in a total number of 540 kinase-substrates possible associations between 82 unique kinases and 267 unique substrates (Table S4). Then, each phosphosite abundance was modeled as a function of targetable kinases via a multivariate linear regression adjusting for treatment information. When both phospho-abundance and global-abundance data were available for a particular kinase, the data type with higher correlation with the targeted site was considered

in the model. For each diagnosed subtype, we adjusted for multiple comparisons via permutation technique. In particular, for each permutation, we run the multivariate analysis after randomly permuting the sample order of the abundance of the targeted site. Repeating this analysis for 200 permutations, we generated the distribution of p-values under the null hypothesis of no-association and utilized this distribution to compute FDR (Tusher et al., 2001). Only associations passing an FDR adjustment of 10% were reported as significant (Table S4, Fig. S4A). Note, given the small sample size of the ATRT and MB cohorts, their shared identity as embryonal tumors and their proteomic similarity (Fig. 1D), ATRT and MB samples were combined to form the ATRT/MB group in this analysis.

**Validation of kinase-phospho associations—**Kinase-phospho associations detected in HGG were validated using proteomic and phosphoproteomic data for 23 additional high-grade glioma samples. This additional data is reported in Table S4.

**BRAF mutation association analysis for LGG—**A regression analysis was performed to compare the abundance of wild type to each mutant type (fusion or point), with age of diagnosis and diagnosis type (initial or progressed) as covariates (Table S5). For this analysis, the LGG diagnosis-specific imputed tables (N=93) were used, including of 5629 and 3437 markers for protein and phosphoproteomic data respectively. Association analysis was also performed for 85 LGG samples across 18209 transcripts.

**Pathways associated with BRAF status in LGG—**Wilcoxon enrichment analysis (WEA) was used to test for association between pathway genesets and *BRAF* status among LGG samples based on regression results from RNA and protein data. Gene set enrichment was conducted across multiple collections of genesets, including GO (Ashburner et al., 2000), Biocarta, KEGG (Kanehisa et al., 2017), Hallmark (Liberzon et al., 2015), and Reactome (Fabregat et al., 2018). These collections were downloaded from http://software.broadinstitute.org/gsea/msigdb/index.jsp on 2/14/2019. Gene sets with less than 5 or more than 250 member genes were excluded. A total of 4795 and 6215 genesets fitting this criterion were tested for enrichment in proteomic and RNA datasets respectively.

<u>**Consolidation pathways via Sumer:**</u> To help identify pathways distinctly associated with each mutation type and to consolidate redundant pathway results, Sumer software was utilized (Savage et al., 2019). The -log10 signed p-value derived from a Z-test comparing mean ssGSEA scores (Hanzelmann et al., 2013) between mutant types was used as the pathway weight when running Sumer. Consolidated pathway modules are shown based on the top 150 pathways by weight (Table S5); a subset of these along with other pathways with biological relevance in cancer were selected for discussion in the main text.

**Phosphoproteomic Co-expression Network in LGG—**Network inference was utilized to characterize co-expression patterns among phosphorylation sites in LGG. The co-expression network was estimated based on phosphosite level data through a random-forest based algorithm (Petralia et al., 2016; Petralia et al., 2015). In particular, co-expression networks were estimated using LGG-specific imputed phosphorylation data. In order to deal with the fact that sites mapping to the same protein are usually correlated, we only modeled each site as function of sites mapping to other proteins. Let $p$ be the total number of sites

measured for $n$ samples. Specifically, let $x_{i,j}^s$ be the abundance of the $j$-th site mapping to the $s$-th protein for the $i$-th sample. Then, $x_{i,j}^s$ was modeled as a function of other protein phosphosites, i.e. $\left\{x_{i,j}^k\right\}_{k \neq s}$, via random forest. In order to derive the final unweighted networks, a proper cut-off value was chosen via permutation techniques (Fruchterman and Reingold, 1991; Petralia et al., 2016). Specifically, 50 permutations and an FDR cut-off of 1E-4 was considered to derive the final network (Table S5). For the visualization of network modules (Fig. 5C, S5C) the software iCAVE (Kalayci and Gumus, 2018; Liuashvili et al., 2017) and Cytoscape (Shannon et al., 2003) were utilized. Force-directed layout algorithm (Fruchterman and Reingold, 1991) was applied to calculate initial positioning of nodes, node positions were then manually adjusted for visual concerns.

**Network modules associated with BRAF status**—Based on the network topology, network modules were identified using an algorithm based on edge betweenness score (Csárdi and Nepusz, 2006; Newman and Girvan, 2004). A total number of 18 network modules containing more than 20 phosphosites were derived (Table S5). Given a network module, the association with $BRAF^{V600E}$ and $BRAF^{Fusion}$ was found via fisher-exact test. In particular, a one-sided fisher exact test was performed to find modules enriched of sites differentially expressed between $BRAF^{Fusion}$ and $BRAF$ wild-type and $BRAF^{V600E}$ and $BRAF$ wild-type at 10% FDR (Table S5). P-values were then adjusted for multiple comparison via Benjamini-Hochberg adjustment (Benjamini and Hochberg, 1995).

**Network modules and druggable kinases**—For this analysis, we considered kinases, which have been used, in clinical trials based on Open Targets database (https://www.targetvalidation.org/disease/EFO_0000311) (Koscielny et al., 2017). A total number of 52 druggable kinases were observed in global proteomic data based on LGG-specific imputed global proteomics table. The association between the global abundance of each kinase and phospho-abundance of phosphosite was assessed via a linear regression. P-values were adjusted for multiple comparison via Benjamini-Hochberg adjustment (Benjamini and Hochberg, 1995) and only associations passing a 5% FDR were reported as significant. Then, to assess the enrichment of sites positively associated to a particular module a one-sided fisher-exact test was performed (Table S5).

**Pathway analysis of network modules**—Gene level pathway analysis was performed for network Module 1 and 2 (referred to as Cluster 1 and 4, respectively in Table S5). Basically, for each network module, we considered the genes whose phosphosites were contained in the network module and identified pathways in the Kegg, Reactome, Hallmark and GO databases enriched in this list. Specifically, a one-sided fisher exact test was performed. Only pathways containing at least 20 genes with phosphorylation measurement were considered for this analysis. Pathways significantly enriched at 10% FDR were found only for Module 1 (Table S5).

**Survival analysis of HGG**—For this analysis, diagnosis-specific imputed proteomic data was used. There were 25 HGG samples which included 3 patients who had two tumor samples at different time points. For these 3 patients, we used the sample of the initial CNS tumor or the one with the smaller age at specimen diagnosis (if both tumors from the same

patient were labelled as progressive/recurrent). Furthermore, 3 HGG samples of autopsies were removed from the analysis. This filtering resulted in 19 HGG samples which were utilized for the survival analysis. Out of these 19 samples, 7 were H3 mutants (all deceased); while the remaining 12 patients (4 alive and 8 deceased) were H3 wild-type. Note that the overall survival was truncated at 2000 days (roughly 5 years) and we treated patients with survival time longer than 2000 days as censored. In particular, only one sample had overall survival greater than 2000 days. This totaled 5 censored samples and 7 "deceased" samples in the H3 wild type group. Considering these samples, survival data was modeled via Cox regression as follows:

$$\text{Coxph(OS, status)} \sim H3^{Mut} + \text{age} + \text{gender} + \text{post\_treatment} + \text{tumor\_purity} + \text{prot} * H3^{Mut} + \text{prot} * H3^{WT} \quad (*)$$

with status denoting the overall survival status. H3_mutation was coded as one for H3 mutant and zero for wild-type. Gender was coded as 1 for male and 0 for female. Treatment status was coded as 1 for "post-treatment" samples and 0 otherwise. The last two terms in the model denote the interaction between protein abundance and H3 mutant and H3 wild-type, respectively. In particular, H3_WT was coded as 1 for H3 wild-type and 0 otherwise; while H3_mut was coded as 1 for H3 mutant and 0 otherwise. Given that for some HGG tumors, transcriptomic data was not available, for this analysis tumor purity was derived based on global proteomic data via TSNet (Petralia et al., 2018) (Table S2).

For 18 of the 19 samples, gene expression data was measured, and we performed a parallel Cox regression analysis based on RNA expression of IDH genes. We derived a 90% confidence interval of hazard ratio estimates for IDH1/2/3 genes based on both global proteomic and gene expression data (Figs. S6A, S6B).

To obtain the effect of IDH1 and IDH2 on survival in the H3 wild type group, we modeled the survival data as function of both IDH1 and IDH2 expression conditional on other covariates as follows:

$$\text{Coxph(OS, status)} \sim H3^{Mut} + \text{age} + \text{gender} + \text{post\_treatment} + \text{tumor\_purity} + IDH1_{pro} * H3^{mut} + IDH2_{pro} * H3^{mut} + IDH1_{pro} * H3^{WT} + IDH2_{pro} * H3^{WT}$$

For assessing the association between the joint effect of IDH1 and IDH2 proteins on overall survival in the H3 wild type group, we performed an anova test to compare the above Cox model with the following one:

$$\text{Model 0: Coxph(OS, status)} \sim H3^{Mut} + \text{age} + \text{gender} + \text{post\_treatment} + \text{tumor\_purity} + IDH1_{pro} * H3^{mut} + IDH2_{pro} * H3^{mut}$$

Let the absolute value of the estimated coefficients of $IDH1_{pro}*H3^{mut}$ and $IDH1_{pro}*H3^{WT}$ be $m_1$ and $w_1$ while those for $IDH2_{pro}*H3^{mut}$ and $IDH2_{pro}*H3^{WT}$ be $m_1$ and $w_1$. We calculate the weighted score of IDH1 and IDH2 for the mutant samples as

$$\text{IDH1/2}^{\text{mut}}{}_{\text{pro}} * \text{H3}^{\text{mut}} = \frac{m_1}{m_1 + m_2} * \text{IDH1}_{\text{pro}} * \text{H3}^{\text{mut}} + \frac{m_2}{m_1 + m_2} * \text{IDH2}_{\text{pro}} * \text{H3}^{\text{mut}}$$

and similarly for the H3 wild type samples as

$$\text{IDH1/2}^{\text{WT}}{}_{\text{pro}} * \text{H3}^{\text{WT}} = \frac{w_1}{w_1 + w_2} * \text{IDH1}_{\text{pro}} * \text{H3}^{\text{WT}} + \frac{w_2}{w_1 + w_2} * \text{IDH2}_{\text{pro}} * \text{H3}^{\text{WT}}$$

In order to display the association between survival and weighted IDH1/2 scores in the H3 wild type group, Kaplan-Meier curves were derived based on $\text{IDH1/2}^{\text{WT}}{}_{\text{pro}}$ (Hanzelmann et al., 2013); with median value chosen as the cut-off to stratify samples in higher and lower abundance groups (Fig. 6E).

We also performed Cox regression to evaluate the association between weighted score of IDH1 and IDH2 with survival conditional on other covariates as follows:

$$\text{Coxph}(\text{OS}, \text{status}) \sim \text{H3\_mutation} + \text{age} + \text{gender} + \text{post\_treatment} + \text{tumor\_purity} + \text{IDH1/2}^{\text{mut}}{}_{\text{pro}} * \text{H3}^{\text{mut}}$$
$$+ \text{IDH1/2}^{\text{WT}}{}_{\text{pro}} * \text{H3}^{\text{WT}}$$

We derived 95% confidence interval of the Hazard ratio estimate and other covariates based on the above model (Fig. 6C).

We also assessed the association between wild type IDH1/2 proteins with survival using a second proteomic dataset containing 41 pediatric and young adult HGG patients without IDH1/2 mutants. Among the 41 samples, 12 samples were H3 mutant: 2 alive and 10 deceased. And the remaining 29 samples (19 deceased and 10 alive) were H3 wild-type (Table S6). For survival analysis, we truncated the OS at 2000 days and treat samples with OS greater than 2000 days as censored samples. This left 17 samples with the deceased status in the H3 wild type group. Similar to the discovery dataset, Cox regression models were fitted on this second data set, with tumor location further included as a covariate. We used indicators for "cortical" and "midline" tumor location, while cerebellum was taken as the reference. Given that only few markers were available for this dataset, we were unable to derive tumor purity and include it as a covariate in the model. The KM curve based on weighted IDH1/2 score is displayed in Figure 6F and the 95% confidence interval of the hazard ratio of weighted score in H3 wild type and other covariates is displayed in Figure 6D.

For pathway enrichment analysis in the pediatric cohort, we used the canonical and Hallmark database from Broad Institute's molecular signature database (Liberzon et al., 2015). We performed a Wilcoxon test to compare the distribution of signed p-values (from Cox regression analysis) of the genes within the pathways to the remaining genes in the dataset. We further consolidated the pathways into modules using Sumer (Savage et al., 2019) (Fig. S6C). Note that we only report the pathway enrichment results from HGG wild

type as this group has reasonably higher sample size as opposed to the mutant group (Table S6).

**Drug Connectivity Analysis for HGG—**For the transcriptional connectivity analysis, an HGG-specific signature was first generated by comparing the mRNA levels between HGG and LGG samples using the Wilcoxon rank sum test. Genes with an FDR < 0.05 were considered differentially expressed and were subsequently filtered for probes measured in the L1000 assay (Subramanian et al., 2017). The resulting gene list was then used as input for iLINCS, a drug connectivity tool (Pilarczyk et al., 2019) and the "Perturbagen connectivity analysis" functionality was used to identify compounds with negative connectivity to the HGG-specific signature.

For the phosphoproteomic connectivity analysis, protein and phosphopeptide signatures were calculated by comparing HGG and LGG samples via the Wilcoxon rank sum test. Significant phosphopeptide and protein probes (FDR < 0.05) were then mapped to the P100 peptide probes (Litichevskiy et al., 2018) and were used for subsequent analysis. Level 4 P100 data were downloaded from the LINCS Data Portal (Stathias et al., 2019) and the median of each technical replicate was used to calculate the spearman correlation between each P100 experiment and the HGG-specific phosphoproteomic signature. The resulting connectivity scores were then aggregated to the compound level, by calculating the mean among all 7 P100 cell lines. To identify drug MOAs (Mechanisms of Action) that were enriched in the transcriptional and phosphoproteomic connectivity analysis we utilized the fgsea R package (https://www.biorxiv.org/content/10.1101/060012v2) by querying against MOA drug sets rather than gene sets. Results for multiple data types are included in Table S6.

**Comparison of initial and progressed tumors—**Our dataset contained proteogenomic profiles of 18 pairs of samples from the same patients. Out of these 18 pairs, 13 of the primary tumors were from the initial disease occurrence, while in the remaining five cases the primary available sample was from a disease that is already classified as progression. Among the secondary samples, 11 are classified as disease progression and seven as recurrence. We analyzed the 18 pairs as cases of less advanced vs more advanced disease, and usually refer to them as initial vs recurrent samples. The mutations included in overlap analysis were the non-synonymous mutations in protein coding genes. Potential driver mutations were either genes with known roles in cancer (Bailey et al., 2018; Li et al., 2015; Li et al., 2018b) that were found to be mutated, or the ones whose allele frequency increased sufficiently between initial and recurrent samples to indicate the signs of evolutionary selection (Merlo et al., 2006). Chromosome arm copy number activity was defined by comparing arm-level CNVs with their standard deviation across all samples. In particular, arm level amplification was declared if arm-level CNVs were over 1-fold SD above zero, while deletion if arm-level CNVs were more than 1-fold SD lower than zero.

Pathway score differences were computed by subtracting the ssGSEA score (Hanzelmann et al., 2013) of a given pathway in the initial sample from that of the recurrent sample.

**Germline variants in TP53—**To screen for germline *TP53* variants that are likely to be pathogenic to or causing Li-Fraunemi syndrome, we checked WGS data fromblood/normal samples in CBTTC (n=893). After filtering, we kept germline variants that were either reported before within Li-fraumeni syndrome patients in the literature according to professional version 2019Q2 of the Human Gene Mutation Database (HGMD)® (Stenson et al., 2017), or predicted to be deleterious in *TP53*, which are defined to be i) in the exonic/ splicing region, ii) not synonoymous SNVs, and iii) with minor allele frequency <0.001 in both of the gnomAD exome and genome databases (version 2.1.1) (Karczewski et al., 2019). Finally we obtained 19 *TP53* variants in 19 CBTTC patients' germline WGS data.

## ADDITIONAL RESOURCES

**Heatmap Web Server—**We have developed a web application (http://pbt.cptac-data-view.org/) which allows researchers to render interactive heatmaps of genes of interest across the cohort, allowing deeper exploration of trends among multiomic and clinical data. The underlying data consists of quantitative information on mutation status, protein abundance, RNA-Seq gene expression, copy number variation, and phosphosite expression for 218 samples when available. The portal has several views available, depending on the data types that the user would like to explore. These views include "all", "mutation", "rna", "proteo", "cnv", and "phospho".

The "all" view provides a multiomic view across multiple data types. Data tracks for each gene are labeled with the gene symbol followed by: "mut" -- ("Yes" for any type of mutation, "No" for wild type), "rna" -- standardized gene expression levels, or "proteo" -- standardized gene-level protein abundance. The "mutation", "rna", "proteo", "cnv", and "phospho" views visualize the individual data tracks. The "phospho" view appends the gene name with a truncated identifier with the amino acid location of the phosphosite, and the user can click the track to see the entire phosphosite identifier.

All views display the genomic and clinical annotation data as the top tracks. These tracks include survival status, grade, diagnosis, tumor location, and clustering analysis results, including an immune cluster assignment for each sample. The genomic annotation tracks include the mutation status for key genes, including *BRAF* status for LGG, *RELA* for EP, *CTNNB1* for CP, and H3F3A for HGG .

The views can show data for the samples across all histological diagnoses or one individual diagnosis (i.e. EP, MB, ATRT, CP, HGG, ganglioglioma, and LGG).

The application can be accessed with any modern web browser through the following address: http://pbt.cptac-data-view.org. Users begin with a text field, where they can enter gene symbols for up to 30 genes. The genes will be used to generate an Excel file (.xls) and heatmap visualizations across all of the views.

Users can click any point on the interactive heatmap to view the underlying data, including sample identifier, data type, and value. They can then sort the heatmap by a given data track, in ascending or descending order. The sorting feature allows researchers to dynamically explore relationships and patterns among various molecular and clinical data types.

## RESOURCE AVAILABILITY

**Lead Contact**—Further information and requests for resources should be directed to and will be fulfilled by the Lead Contact, Pei Wang (pei.wang@mssm.edu).

**Materials Availability**—N/A

**Data and Code Availability**—All raw genomic data is available upon access request through the Children's Brain Tumor Tissue Consortium (https://cbttc.org/) and can be accessed through the Gabriella Miller Kids First Portal (https://kidsfirstdrc.org/). All raw proteomics data and processed proteogenomic data are available through the Clinical Proteomic Tumor Analysis Consortium Data Portal (https://cptac-data-portal.georgetown.edu/cptacPublic/) and the Proteomics Data Commons (https://pdc.cancer.gov/pdc/). In addition, all processed proteogenomic data sets as well as clinical meta information can be queried, visualized and downloaded from an interactive *ProTrack* data portal http://pbt.cptac-data-view.org/, as well as through the *PedcBioPortal* (https://pedcbioportal.kidsfirstdrc.org/).

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Authors

Francesca Petralia[1,34], Nicole Tignor[1,34], Boris Reva[1,34], Mateusz Koptyra[2,34], Shrabanti Chowdhury[1,34], Dmitry Rykunov[1,35], Azra Krek[1,35], Weiping Ma[1,35], Yuankun Zhu[2,35], Jiayi Ji[3,4], Anna Calinawan[1], Jeffrey R. Whiteaker[5], Antonio Colaprico[6], Vasileios Stathias[8], Tatiana Omelchenko[9], Xiaoyu Song[3,4], Pichai Raman[2,10,32], Yiran Guo[2], Miguel A. Brown[2], Richard G. Ivey[5], John Szpyt[11], Sanjukta Guha Thakurta[11], Marina A. Gritsenko[12], Karl K. Weitz[12], Gonzalo Lopez[1], Selim Kalayci[1], Zeynep H. Gümü[1], Seungyeul Yoo[1], Felipe da Veiga Leprevost[13], Hui-Yin Chang[13], Karsten Krug[14], Lizabeth Katsnelson[15], Ying Wang[15], Jacob J. Kennedy[5], Uliana J. Voytovich[5], Lei Zhao[5], Krutika S. Gaonkar[2,10,32], Brian M. Ennis[2], Bo Zhang[2], Valerie Baubet[2], Lamiya Tauhid[2], Jena V. Lilly[2], Jennifer L. Mason[2], Bailey Farrow[2], Nathan Young[2], Sarah Leary[5,16,17], Jamie Moon[12], Vladislav A. Petyuk[12], Javad Nazarian[18,19], Nithin D. Adappa[20], James N. Palmer[20], Robert M. Lober[21], Samuel Rivero-Hinojosa[18], Liang-Bo Wang[22], Joshua M. Wang[15], Matilda Broberg[15], Rosalie K. Chu[12], Ronald J. Moore[12], Matthew E. Monroe[12], Rui Zhao[12], Richard D. Smith[12], Jun Zhu[1], Ana I. Robles[24], Mehdi Mesri[24], Emily Boja[24], Tara Hiltke[24], Henry Rodriguez[24], Bing Zhang[25,26,27], Eric E. Schadt[1], D. R. Mani[14], Li Ding[22,23], Antonio Iavarone[28], Maciej Wiznerowicz[29,30], Stephan Schürer[8], Xi S. Chen[6,7], Allison P. Heath[2], Jo Lynne Rokita[2,10,32], Alexey I. Nesvizhskii[13,31], David Fenyö[15], Karin D. Rodland[12,33], Tao Liu[12], Steven P. Gygi[11,36], Amanda G. Paulovich[5,36], Adam C Resnick[2,32,*], Phillip B. Storm[2,32,*], Brian R. Rood[18,*], Pei Wang[1,37,*], Children's Brain Tumor Network, Clinical Proteomic Tumor Analysis Consortium

## Affiliations

[1]Department of Genetics and Genomic Sciences and Icahn Institute for Data Science and Genomic Technology, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA [2]Center for Data-Driven Discovery in Biomedicine, Division of Neurosurgery, Children's Hospital of Philadelphia, Philadelphia, PA, 19104, USA [3]Department of Population Health Science and Policy, Icahn School of Medicine at Mount Sinai, New York, NY, 10029, USA [4]Tisch Cancer Institute, Icahn School of Medicine at Mount Sinai, New York, NY, 10029, USA [5]Fred Hutchinson Cancer Research Center, Seattle, WA, 98109, USA [6]Department of Public Health Science, University of Miami Miller School of Medicine, Miami, FL, 33136,USA [7]Sylvester Comprehensive Cancer Center, University of Miami Miller School of Medicine, Miami, FL, 33136,USA [8]Department of Pharmacology, Institute for Data Science and Computing, Sylvester Comprehensive Cancer Center, University of Miami, Miami, FL, 33146, USA [9]Cell Biology Program, Memorial Sloan Kettering Cancer Center, New York, NY, 10065, USA [10]Department of Bioinformatics and Health Informatics, Children's Hospital of Philadelphia, Philadelphia, PA, 19104, USA [11]Thermo Fisher Scientific Center for Multiplexed Proteomics, Department of Cell Biology, Harvard Medical School, Boston, MA, 02115, USA [12]Biological Sciences Division, Pacific Northwest National Laboratory, Richland, WA, 99354, USA [13]Department of Pathology, University of Michigan, Ann Arbor, MI, 48109, USA [14]Broad Institute of Massachusetts Institute of Technology and Harvard, Cambridge, MA, 02412, USA [15]Institute for Systems Genetics; Department of Biochemistry and Molecular Pharmacology, NYU Grossman School of Medicine, New York, NY, 10016, USA [16]Cancer and Blood Disorders Center, Seattle Children's Hospital, Seattle, 98105, WA [17]Department of Pediatrics, University of Washington, Seattle, 98195, WA [18]Children's National Research Institute, Washington DC, 20010, USA [19]Department of Oncology, Children's Research Center, University Children's Hospital Zürich, Zurich, 8032, Switzerland [20]Department of Otorhinolaryngology, University of Pennsylvania, Philadelphia, PA, 19104, USA [21]Department of Neurosurgery, Dayton Children's Hospital and Wright State University Boonshoft School of Medicine, Dayton, OH, 45404, USA [22]Department of Medicine, Washington University in St. Louis, St. Louis, MO 631110, USA; McDonnell Genome Institute, Washington University in St. Louis, St. Louis, MO 63108, USA [23]Siteman Cancer Center, Washington University in St. Louis, St. Louis, MO 63110, USA [24]Office of Cancer Clinical Proteomics Research, National Cancer Institute, National Institutes of Health, Bethesda, MD, 20892, USA [25]Lester and Sue Smith Breast Center, Baylor College of Medicine, Houston, TX 77030, USA [26]Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030, USA [27]Dan L Duncan Comprehensive Cancer Center, Baylor College of Medicine, Houston, TX 77030, USA [28]Institute for Cancer Genetics, Department of Neurology, Department of Pathology and Cell Biology, Herbert Irving Comprehensive Cancer Center, Columbia University Medical Center, New York, NY 10032, USA [29]Poznan University of Medical Sciences, 61-701 Poznan, Poland [30]International Institute for Molecular Oncology, 61-203 Pozna, Poland [31]Department of Computational

Medicine & Bioinformatics, University of Michigan, Ann Arbor, MI, 48109, USA [32]Division of Neurosurgery, Children's Hospital of Philadelphia, Philadelphia, PA, 19104, USA [33]Department of Cell, Developmental, and Cancer Biology, Oregon Health & Science University, Portland, OR 97221, USA [34]These authors contributed equally [35]These authors contributed equally [36]Senior authors [37]Lead contact: pei.wang@mssm.edu (P.W.)

## Acknowledgement

## REFERENCES

Abate M, Laezza C, Pisanti S, Torelli G, Seneca V, Catapano G, Montella F, Ranieri R, Notarnicola M, Gazzerro P, et al. (2017). Deregulated expression and activity of Farnesyl Diphosphate Synthase (FDPS) in Glioblastoma. Scientific reports 7, 14123. [PubMed: 29075041]

Apps JR, Carreno G, Gonzalez-Meljem JM, Haston S, Guiho R, Cooper JE, Manshaei S, Jani N, Holsken A, Pettorini B, et al. (2018). Tumour compartment transcriptomics demonstrates the activation of inflammatory and odontogenic programmes in human adamantinomatous craniopharyngioma and identifies the MAPK/ERK pathway as a novel therapeutic target. Acta neuropathologica 135, 757–777. [PubMed: 29541918]

Aran D, Hu Z, and Butte AJ (2017). xCell: digitally portraying the tissue cellular heterogeneity landscape. Genome biology 18, 220. [PubMed: 29141660]

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, et al. (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nature genetics 25, 25–29. [PubMed: 10802651]

Bailey MH, Tokheim C, Porta-Pardo E, Sengupta S, Bertrand D, Weerasinghe A, Colaprico A, Wendl MC, Kim J, Reardon B, et al. (2018). Comprehensive Characterization of Cancer Driver Genes and Mutations. Cell 173, 371–385.e318. [PubMed: 29625053]

Baroti T, Zimmermann Y, Schillinger A, Liu L, Lommes P, Wegner M, and Stolt CC (2016). Transcription factors Sox5 and Sox6 exert direct and indirect influences on oligodendroglial migration in spinal cord and forebrain. Glia 64, 122–138. [PubMed: 26345464]

Beausoleil SA, Villén J, Gerber SA, Rush J, and Gygi SP (2006). A probability-based approach for high-throughput protein phosphorylation analysis and site localization. Nature biotechnology 24, 1285–1292.

Bednarek E, and Caroni P (2011). beta-Adducin is required for stable assembly of new synapses and improved memory upon environmental enrichment. Neuron 69, 1132–1146. [PubMed: 21435558]

Behrens J, Kameritsch P, Wallner S, Pohl U, and Pogoda K (2010). The carboxyl tail of Cx43 augments p38 mediated cell migration in a gap junction-independent manner. European journal of cell biology 89, 828–838. [PubMed: 20727616]

Belletti B, and Baldassarre G (2011). Stathmin: a protein with many tasks. New biomarker and potential target in cancer. Expert opinion on therapeutic targets 15, 1249–1266. [PubMed: 21978024]

Benjamini Y, and Hochberg Y (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society Series B (Methodological) 57, 289–300.

Bergaggio E, and Piva R (2019). Wild-Type IDH Enzymes as Actionable Targets for Cancer Therapy. Cancers 11.

Box JK, Paquet N, Adams MN, Boucher D, Bolderson E, O'Byrne KJ, and Richard DJ (2016). Nucleophosmin: from structure and function to disease development. BMC molecular biology 17, 19. [PubMed: 27553022]

Brastianos PK, Taylor-Weiner A, Manley PE, Jones RT, Dias-Santagata D, Thorner AR, Lawrence MS, Rodriguez FJ, Bernardo LA, Schubert L, et al. (2014). Exome sequencing identifies BRAF mutations in papillary craniopharyngiomas. Nature genetics 46, 161–165. [PubMed: 24413733]

Briancon-Marjollet A, Balenci L, Fernandez M, Esteve F, Honnorat J, Farion R, Beaumont M, Barbier E, Remy C, and Baudier J (2010). NG2-expressing glial precursor cells are a new potential oligodendroglioma cell initiating population in N-ethyl-N-nitrosourea-induced gliomagenesis. Carcinogenesis 31, 1718–1725. [PubMed: 20651032]

Buckingham SC, Campbell SL, Haas BR, Montana V, Robel S, Ogunrinu T, and Sontheimer H (2011). Glutamate release by primary brain tumors induces epileptic activity. Nature medicine 17, 1269–1274.

Butovsky O, Jedrychowski MP, Moore CS, Cialic R, Lanser AJ, Gabriely G, Koeglsperger T, Dake B, Wu PM, Doykan CE, et al. (2014). Identification of a unique TGF-beta-dependent molecular and functional signature in microglia. Nature neuroscience 17, 131–143. [PubMed: 24316888]

Cai Y, Guo T, Wang Y, and Du J (2018). Glutamate Metabolism Regulates Immune Escape of Glioma. Madridge Journal of Immunology 2, 53–57.

Cai Z, Wang Y, Yu W, Xiao J, Li Y, Liu L, Zhu C, Tan K, Deng Y, Yuan W, et al. (2006). hnulp1, a basic helix-loop-helix protein with a novel transcriptional repressive domain, inhibits transcriptional activity of serum response factor. Biochemical and biophysical research communications 343, 973–981. [PubMed: 16574069]

Calvert AE, Chalastanis A, Wu Y, Hurley LA, Kouri FM, Bi Y, Kachman M, May JL, Bartom E, Hua Y, et al. (2017). Cancer-Associated IDH1 Promotes Growth and Resistance to Targeted Therapies in the Absence of Mutation. Cell reports 19, 1858–1873. [PubMed: 28564604]

Campanini ML, Colli LM, Paixao BM, Cabral TP, Amaral FC, Machado HR, Neder LS, Saggioro F, Moreira AC, Antonini SR, et al. (2010). CTNNB1 gene mutations, pituitary transcription factors, and MicroRNA expression involvement in the pathogenesis of adamantinomatous craniopharyngiomas. Hormones & cancer 1, 187–196. [PubMed: 21761366]

Campbell SL, Buckingham SC, and Sontheimer H (2012). Human glioma cells induce hyperexcitability in cortical networks. Epilepsia 53, 1360–1370. [PubMed: 22709330]

Campbell SL, Robel S, Cuddapah VA, Robert S, Buckingham SC, Kahle KT, and Sontheimer H (2015). GABAergic disinhibition and impaired KCC2 cotransporter activity underlie tumor-associated epilepsy. Glia 63, 23–36. [PubMed: 25066727]

Cancer Genome Atlas Research, N. (2008). Comprehensive genomic characterization defines human glioblastoma genes and core pathways. Nature 455, 1061–1068. [PubMed: 18772890]

Cancer Genome Atlas Research, N., Brat DJ, Verhaak RG, Aldape KD, Yung WK, Salama SR, Cooper LA, Rheinbay E, Miller CR, Vitucci M, et al. (2015). Comprehensive, Integrative Genomic Analysis of Diffuse Lower-Grade Gliomas. N Engl J Med 372, 2481–2498. [PubMed: 26061751]

Capper D, Jones DTW, Sill M, Hovestadt V, Schrimpf D, Sturm D, Koelsche C, Sahm F, Chavez L, Reuss DE, et al. (2018). DNA methylation-based classification of central nervous system tumours. Nature 555, 469–474. [PubMed: 29539639]

Cesca F, Baldelli P, Valtorta F, and Benfenati F (2010). The synapsins: key actors of synapse function and plasticity. Progress in neurobiology 91, 313–348. [PubMed: 20438797]

Chalmers ZR, Connelly CF, Fabrizio D, Gay L, Ali SM, Ennis R, Schrock A, Campbell B, Shlien A, Chmielecki J, et al. (2017). Analysis of 100,000 human cancer genomes reveals the landscape of tumor mutational burden. Genome medicine 9, 34. [PubMed: 28420421]

Chang CV, Araujo RV, Cirqueira CS, Cani CM, Matushita H, Cescato VA, Fragoso MC, Bronstein MD, Zerbini MC, Mendonca BB, et al. (2017). Differential Expression of Stem Cell Markers in Human Adamantinomatous Craniopharyngioma and Pituitary Adenoma. Neuroendocrinology 104, 183–193. [PubMed: 27161333]

Chen JH, Huang SM, Chen CC, Tsai CF, Yeh WL, Chou SJ, Hsieh WT, and Lu DY (2011). Ghrelin induces cell migration through GHS-R, CaMKII, AMPK, and NF-κB signaling pathway in glioma cells. Journal of cellular biochemistry 112, 2931–2941. [PubMed: 21630326]

Chisci E, De Giorgi M, Zanfrini E, Testasecca A, Brambilla E, Cinti A, Farina L, Kutryb-Zajac B, Bugarin C, Villa C, et al. (2017). Simultaneous overexpression of human E5NT and ENTPD1 protects porcine endothelial cells against H2O2-induced oxidative stress and cytotoxicity in vitro. Free radical biology & medicine 108, 320–333. [PubMed: 28389406]

Clark DJ, Dhanasekaran SM, Petralia F, Pan J, Song X, Hu Y, da Veiga Leprevost F, Reva B, Lih TM, Chang HY, et al. (2019). Integrated Proteogenomic Characterization of Clear Cell Renal Cell Carcinoma. Cell 179, 964–983.e931. [PubMed: 31675502]

Colaprico A, Olsen C, Bailey MH, Odom GJ, Terkelsen T, Silva TC, Olsen AV, Cantini L, Zinovyev A, Barillot E, et al. (2020). Interpreting pathways to discover cancer driver genes with Moonlight. Nature communications 11, 69.

Colaprico A, Silva TC, Olsen C, Garofano L, Cava C, Garolini D, Sabedot TS, Malta TM, Pagnotta SM, Castiglioni I, et al. (2016). TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. Nucleic acids research 44, e71. [PubMed: 26704973]

Cole AR (2012). GSK3 as a Sensor Determining Cell Fate in the Brain. Frontiers in molecular neuroscience 5, 4. [PubMed: 22363258]

Cooper CD, and Lampe PD (2002). Casein kinase 1 regulates connexin-43 gap junction assembly. The Journal of biological chemistry 277, 44962–44968. [PubMed: 12270943]

Coy S, Rashid R, Lin JR, Du Z, Donson AM, Hankinson TC, Foreman NK, Manley PE, Kieran MW, Reardon DA, et al. (2018). Multiplexed immunofluorescence reveals potential PD-1/PD-L1 pathway vulnerabilities in craniopharyngioma. Neuro-oncology 20, 1101–1112. [PubMed: 29509940]

Crotti A, and Ransohoff RM (2016). Microglial Physiology and Pathophysiology: Insights from Genome-wide Transcriptional Profiling. Immunity 44, 505–515. [PubMed: 26982357]

Csárdi G, and Nepusz T (2006). The igraph software package for complex network research.

Cuddapah VA, and Sontheimer H (2010). Molecular interaction and functional regulation of ClC-3 by Ca2+/calmodulin-dependent protein kinase II (CaMKII) in human malignant glioma. The Journal of biological chemistry 285, 11188–11196. [PubMed: 20139089]

da Veiga Leprevost F, Haynes SE, Avtonomov DM, Chang H, Shanmugam AK, Mellacheruvu D, Kong AT, and Nesvizhskii AI (2020). Philosopher: a versatile toolkit for shotgun proteomics data analysis. Nature methods in press.

Daily K, Ho Sui SJ, Schriml LM, Dexheimer PJ, 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. Scientific data 4, 170030. [PubMed: 28350385]

Darmanis S, Sloan SA, Croote D, Mignardi M, Chernikova S, Samghababi P, Zhang Y, Neff N, Kowarsky M, Caneda C, et al. (2017). Single-Cell RNA-Seq Analysis of Infiltrating Neoplastic Cells at the Migrating Front of Human Glioblastoma. Cell reports 21, 1399–1410. [PubMed: 29091775]

Dello Russo C, Lisi L, Tentori L, Navarra P, Graziani G, and Combs CK (2017). Exploiting Microglial Functions for the Treatment of Glioblastoma. Current cancer drug targets 17, 267–281. [PubMed: 27528361]

Djomehri SI, Gonzalez ME, da Veiga Leprevost F, Tekula SR, Chang HY, White MJ, Cimino-Mathews A, Burman B, Basrur V, Argani P, et al. (2020). Quantitative proteomic landscape of metaplastic breast carcinoma pathological subtypes and their relationship to triple-negative tumors. Nature communications 11, 1723.

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, and Gingeras TR (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics (Oxford, England) 29, 15–21.

Dou Y, Kawaler EA, Cui Zhou D, Gritsenko MA, Huang C, Blumenberg L, Karpova A, Petyuk VA, Savage SR, Satpathy S, et al. (2020). Proteogenomic Characterization of Endometrial Carcinoma. Cell 180, 729–748.e726. [PubMed: 32059776]

Easley-Neal C, Fierro J Jr., Buchanan J, and Washbourne P (2013). Late recruitment of synapsin to nascent synapses is regulated by Cdk5. Cell reports 3, 1199–1212. [PubMed: 23602570]

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 research 46, D649–d655. [PubMed: 29145629]

Fangusaro J, Onar-Thomas A, Young Poussaint T, Wu S, Ligon AH, Lindeman N, Banerjee A, Packer RJ, Kilburn LB, Goldman S, et al. (2019). Selumetinib in paediatric patients with BRAF-aberrant or neurofibromatosis type 1-associated recurrent, refractory, or progressive low-grade glioma: a multicentre, phase 2 trial. The Lancet Oncology 20, 1011–1022. [PubMed: 31151904]

Farghaian H, Turnley AM, Sutherland C, and Cole AR (2011). Bioinformatic prediction and confirmation of beta-adducin as a novel substrate of glycogen synthase kinase 3. The Journal of biological chemistry 286, 25274–25283. [PubMed: 21606488]

Fasolini M, Wu X, Flocco M, Trosset JY, Oppermann U, and Knapp S (2003). Hot spots in Tcf4 for the interaction with beta-catenin. The Journal of biological chemistry 278, 21092–21098. [PubMed: 12657632]

Fei L, and Xu H (2018). Role of MCM2–7 protein phosphorylation in human cancer cells. Cell & bioscience 8, 43. [PubMed: 30062004]

Finzsch M, Stolt CC, Lommes P, and Wegner M (2008). Sox9 and Sox10 influence survival and migration of oligodendrocyte precursors in the spinal cord by regulating PDGF receptor alpha expression. Development (Cambridge, England) 135, 637–646.

Fruchterman TMJ, and Reingold EM (1991). Graph drawing by force-directed placement. Software: Practice and Experience 21, 1129–1164.

Fumagalli M, Lombardi M, Gressens P, and Verderio C (2018). How to reprogram microglia toward beneficial functions. Glia 66, 2531–2549. [PubMed: 30195261]

Gao C, Wang Y, Broaddus R, Sun L, Xue F, and Zhang W (2018). Exon 3 mutations of CTNNB1 drive tumorigenesis: a review. Oncotarget 9, 5492–5508. [PubMed: 29435196]

Gibbons BC, Chambers MC, Monroe ME, Tabb DL, and Payne SH (2015). Correcting systematic bias and instrument measurement drift with mzRefinery. Bioinformatics (Oxford, England) 31, 3838–3840.

Gillette MA, Satpathy S, Cao S, Dhanasekaran SM, Vasaikar SV, Krug K, Petralia F, Li Y, Liang WW, Reva B, et al. (2020). Proteogenomic Characterization Reveals Therapeutic Vulnerabilities in Lung Adenocarcinoma. Cell 182, 200–225.e235. [PubMed: 32649874]

Grobner SN, Worst BC, Weischenfeldt J, Buchhalter I, Kleinheinz K, Rudneva VA, Johann PD, Balasubramanian GP, Segura-Wang M, Brabetz S, et al. (2018). The landscape of genomic alterations across childhood cancers. Nature 555, 321–327. [PubMed: 29489754]

Grosely R, Kopanic JL, Nabors S, Kieken F, Spagnol G, Al-Mugotir M, Zach S, and Sorgen PL (2013). Effects of phosphorylation on the structure and backbone dynamics of the intrinsically disordered connexin43 C-terminal domain. The Journal of biological chemistry 288, 24857–24870. [PubMed: 23828237]

Haage V, Semtner M, Vidal RO, Hernandez DP, Pong WW, Chen Z, Hambardzumyan D, Magrini V, Ly A, Walker J, et al. (2019). Comprehensive gene expression meta-analysis identifies signature genes that distinguish microglia from peripheral monocytes/macrophages in health and glioma. Acta neuropathologica communications 7, 20. [PubMed: 30764877]

Hanel W, and Moll UM (2012). Links between mutant p53 and genomic instability. Journal of cellular biochemistry 113, 433–439. [PubMed: 22006292]

Hanzelmann S, Castelo R, and Guinney J (2013). GSVA: gene set variation analysis for microarray and RNA-seq data. BMC bioinformatics 14, 7. [PubMed: 23323831]

Hayashi H, Matsuzaki O, Muramatsu S, Tsuchiya Y, Harada T, Suzuki Y, Sugano S, Matsuda A, and Nishida E (2006). Centaurin-alpha1 is a phosphatidylinositol 3-kinase-dependent activator of ERK1/2 mitogen-activated protein kinases. The Journal of biological chemistry 281, 1332–1337. [PubMed: 16287813]

Holcik M (2015). Could the eIF2alpha-Independent Translation Be the Achilles Heel of Cancer? Frontiers in oncology 5, 264. [PubMed: 26636041]

Hong X, Sin WC, Harris AL, and Naus CC (2015). Gap junctions modulate glioma invasion by direct transfer of microRNA. Oncotarget 6, 15566–15577. [PubMed: 25978028]

Hornbeck PV, Zhang B, Murray B, Kornhauser JM, Latham V, and Skrzypek E (2015). PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. Nucleic acids research 43, D512–520. [PubMed: 25514926]

Ito T, Chiba T, Ozawa R, Yoshida M, Hattori M, and Sakaki Y (2001). A comprehensive two-hybrid analysis to explore the yeast protein interactome. Proceedings of the National Academy of Sciences of the United States of America 98, 4569–4574. [PubMed: 11283351]

Itoh T, Satoh M, Kanno E, and Fukuda M (2006). Screening for target Rabs of TBC (Tre-2/Bub2/Cdc16) domain-containing proteins based on their Rab-binding activity. Genes to cells : devoted to molecular & cellular mechanisms 11, 1023–1037. [PubMed: 16923123]

Jain P, Silva A, Han HJ, Lang SS, Zhu Y, Boucher K, Smith TE, Vakil A, Diviney P, Choudhari N, et al. (2017). Overcoming resistance to single-agent therapy for oncogenic BRAF gene fusions via combinatorial targeting of MAPK and PI3K/mTOR signaling pathways. Oncotarget 8, 84697–84713. [PubMed: 29156677]

John Lin CC, Yu K, Hatcher A, Huang TW, Lee HK, Carlson J, Weston MC, Chen F, Zhang Y, Zhu W, et al. (2017). Identification of diverse astrocyte populations and their malignant analogs. Nature neuroscience 20, 396–405. [PubMed: 28166219]

Johnson WE, Li C, and Rabinovic A (2007). Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics (Oxford, England) 8, 118–127.

Kalayci S, and Gumus ZH (2018). Exploring Biological Networks in 3D, Stereoscopic 3D, and Immersive 3D with iCAVE. Current protocols in bioinformatics 61, 8.27.21–28.27.26.

Kanehisa M, Furumichi M, Tanabe M, Sato Y, and Morishima K (2017). KEGG: new perspectives on genomes, pathways, diseases and drugs. Nucleic acids research 45, D353–d361. [PubMed: 27899662]

Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, Collins RL, Laricchia KM, Ganna A, Birnbaum DP, et al. (2019). Variation across 141,456 human exomes and genomes reveals the spectrum of loss-of-function intolerance across human protein-coding genes. bioRxiv, 531210.

Karmakar S, Dey P, Vaz AP, Bhaumik SR, Ponnusamy MP, and Batra SK (2018). PD2/PAF1 at the Crossroads of the Cancer Network. Cancer research 78, 313–319. [PubMed: 29311159]

Karremann M, Gielen GH, Hoffmann M, Wiese M, Colditz N, Warmuth-Metz M, Bison B, Claviez A, van Vuurden DG, von Bueren AO, et al. (2018). Diffuse high-grade gliomas with H3 K27M mutations carry a dismal prognosis independent of tumor location. Neuro-oncology 20, 123–131. [PubMed: 29016894]

Keller A, Nesvizhskii AI, Kolker E, and Aebersold R (2002). Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. Analytical chemistry 74, 5383–5392. [PubMed: 12403597]

Khuong-Quang DA, Buczkowicz P, Rakopoulos P, Liu XY, Fontebasso AM, Bouffet E, Bartels U, Albrecht S, Schwartzentruber J, Letourneau L, et al. (2012). K27M mutation in histone H3.3 defines clinically and biologically distinct subgroups of pediatric diffuse intrinsic pontine gliomas. Acta neuropathologica 124, 439–447. [PubMed: 22661320]

Kilday JP, Mitra B, Domerg C, Ward J, Andreiuolo F, Osteso-Ibanez T, Mauguen A, Varlet P, Le Deley MC, Lowe J, et al. (2012). Copy number gain of 1q25 predicts poor progression-free survival for pediatric intracranial ependymomas and enables patient risk stratification: a prospective European clinical trial cohort analysis on behalf of the Children's Cancer Leukaemia Group (CCLG), Societe Francaise d'Oncologie Pediatrique (SFOP), and International Society for Pediatric

Oncology (SIOP). Clinical cancer research : an official journal of the American Association for Cancer Research 18, 2001–2011. [PubMed: 22338015]

Kim HY, Kim DK, Bae SH, Gwak H, Jeon JH, Kim JK, Lee BI, You HJ, Shin DH, Kim YH, et al. (2018a). Farnesyl diphosphate synthase is important for the maintenance of glioblastoma stemness. Experimental & molecular medicine 50, 137.

Kim S, Gupta N, and Pevzner PA (2008). Spectral probabilities and generating functions of tandem mass spectra: a strike against decoy databases. Journal of proteome research 7, 3354–3363. [PubMed: 18597511]

Kim S, and Pevzner PA (2014). MS-GF+ makes progress towards a universal database search tool for proteomics. Nature communications 5, 5277.

Kim S, Scheffler K, Halpern AL, Bekritsky MA, Noh E, Kallberg M, Chen X, Kim Y, Beyter D, Krusche P, et al. (2018b). Strelka2: fast and accurate calling of germline and somatic variants. Nature methods 15, 591–594. [PubMed: 30013048]

Kim W, and Liau LM (2012). IDH mutations in human glioma. Neurosurgery clinics of North America 23, 471–480. [PubMed: 22748659]

Kong AT, Leprevost FV, Avtonomov DM, Mellacheruvu D, and Nesvizhskii AI (2017). MSFragger: ultrafast and comprehensive peptide identification in mass spectrometry-based proteomics. Nature methods 14, 513–520. [PubMed: 28394336]

Kordass T, Osen W, and Eichmuller SB (2018). Controlling the Immune Suppressor: Transcription Factors and MicroRNAs Regulating CD73/NT5E. Frontiers in immunology 9, 813.

Koschmann C, Zamler D, MacKay A, Robinson D, Wu YM, Doherty R, Marini B, Tran D, Garton H, Muraszko K, et al. (2016). Characterizing and targeting PDGFRA alterations in pediatric high-grade glioma. Oncotarget 7, 65696–65706. [PubMed: 27582545]

Koscielny G, An P, Carvalho-Silva D, Cham JA, Fumis L, Gasparyan R, Hasan S, Karamanis N, Maguire M, Papa E, et al. (2017). Open Targets: a platform for therapeutic target identification and validation. Nucleic acids research 45, D985–d994. [PubMed: 27899665]

Koseki J, Colvin H, Fukusumi T, Nishida N, Konno M, Kawamoto K, Tsunekuni K, Matsui H, Doki Y, Mori M, et al. (2015). Mathematical analysis predicts imbalanced IDH1/2 expression associates with 2-HG-inactivating β-oxygenation pathway in colorectal cancer. International journal of oncology 46, 1181–1191. [PubMed: 25586680]

Kouidhi S, Ben Ayed F, and Benammar Elgaaied A (2018). Targeting Tumor Metabolism: A New Challenge to Improve Immunotherapy. Frontiers in immunology 9, 353. [PubMed: 29527212]

Krasemann S, Madore C, Cialic R, Baufeld C, Calcagno N, El Fatimy R, Beckers L, O'Loughlin E, Xu Y, Fanek Z, et al. (2017). The TREM2-APOE Pathway Drives the Transcriptional Phenotype of Dysfunctional Microglia in Neurodegenerative Diseases. Immunity 47, 566–581.e569. [PubMed: 28930663]

Kunze A, Congreso MR, Hartmann C, Wallraff-Beck A, Hüttmann K, Bedner P, Requardt R, Seifert G, Redecker C, Willecke K, et al. (2009). Connexin expression by radial glia-like cells is required for neurogenesis in the adult dentate gyrus. Proceedings of the National Academy of Sciences of the United States of America 106, 11336–11341. [PubMed: 19549869]

Lakka SS, and Rao JS (2008). Antiangiogenic therapy in brain tumors. Expert review of neurotherapeutics 8, 1457–1473. [PubMed: 18928341]

Lee S, Lee S, Ouellette S, Park WY, Lee EA, and Park PJ (2017). NGSCheckMate: software for validating sample identity in next-generation sequencing studies within and across data types. Nucleic acids research 45, e103. [PubMed: 28369524]

Leone RD, and Emens LA (2018). Targeting adenosine for cancer immunotherapy. Journal for immunotherapy of cancer 6, 57. [PubMed: 29914571]

Leone RD, Zhao L, Englert JM, Sun IM, Oh MH, Sun IH, Arwood ML, Bettencourt IA, Patel CH, Wen J, et al. (2019). Glutamine blockade induces divergent metabolic programs to overcome tumor immune evasion. Science (New York, NY) 366, 1013–1021.

Li B, and Dewey CN (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC bioinformatics 12, 323. [PubMed: 21816040]

Li J, Davidson D, Martins Souza C, Zhong MC, Wu N, Park M, Muller WJ, and Veillette A (2015). Loss of PTPN12 Stimulates Progression of ErbB2-Dependent Breast Cancer by Enhancing Cell

Survival, Migration, and Epithelial-to-Mesenchymal Transition. Molecular and cellular biology 35, 4069–4082. [PubMed: 26391955]

Li J, He Y, Tan Z, Lu J, Li L, Song X, Shi F, Xie L, You S, Luo X, et al. (2018a). Wild-type IDH2 promotes the Warburg effect and tumor growth through HIF1α in lung cancer. Theranostics 8, 4050–4061. [PubMed: 30128035]

Li L, Xing R, Cui J, Li W, and Lu Y (2018b). Investigation of frequent somatic mutations of MTND5 gene in gastric cancer cell lines and tissues. Mitochondrial DNA Part B 3, 1002–1008. [PubMed: 33474396]

Lian CG, Xu Y, Ceol C, Wu F, Larson A, Dresser K, Xu W, Tan L, Hu Y, Zhan Q, et al. (2012). Loss of 5-hydroxymethylcytosine is an epigenetic hallmark of melanoma. Cell 150, 1135–1146. [PubMed: 22980977]

Liberzon A, Birger C, Thorvaldsdottir H, Ghandi M, Mesirov JP, and Tamayo P (2015). The Molecular Signatures Database (MSigDB) hallmark gene set collection. Cell systems 1, 417–425. [PubMed: 26771021]

Liberzon A, Subramanian A, Pinchback R, Thorvaldsdottir H, Tamayo P, and Mesirov JP (2011). Molecular signatures database (MSigDB) 3.0. Bioinformatics (Oxford, England) 27, 1739–1740.

Liluashvili V, Kalayci S, Fluder E, Wilson M, Gabow A, and Gumus ZH (2017). iCAVE: an open source tool for visualizing biomolecular networks in 3D, stereoscopic 3D and immersive 3D. GigaScience 6, 1–13.

Lin AP, Demeler B, Minard KI, Anderson SL, Schirf V, Galaleldeen A, and McAlister-Henn L (2011). Construction and analyses of tetrameric forms of yeast NAD+-specific isocitrate dehydrogenase. Biochemistry 50, 230–239. [PubMed: 21133413]

Lin AP, and McAlister-Henn L (2011). Basis for half-site ligand binding in yeast NAD(+)-specific isocitrate dehydrogenase. Biochemistry 50, 8241–8250. [PubMed: 21861471]

Litichevskiy L, Peckner R, Abelin JG, Asiedu JK, Creech AL, Davis JF, Davison D, Dunning CM, Egertson JD, Egri S, et al. (2018). A Library of Phosphoproteomic and Chromatin Signatures for Characterizing Cellular Responses to Drug Perturbations. Cell systems 6, 424–443.e427. [PubMed: 29655704]

Liu H, Sun Y, O'Brien JA, Franco-Barraza J, Qi X, Yuan H, Jin W, Zhang J, Gu C, Zhao Z, et al. (2019). Necroptotic astrocytes contribute to maintaining stemness of disseminated medulloblastoma through CCL2 secretion. Neuro-oncology.

Liu PK, Kraus E, Wu TA, Strong LC, and Tainsky MA (1996). Analysis of genomic instability in Li-Fraumeni fibroblasts with germline p53 mutations. Oncogene 12, 2267–2278. [PubMed: 8649766]

Liu WR, Tian MX, Jin L, Yang LX, Ding ZB, Shen YH, Peng YF, Zhou J, Qiu SJ, Dai Z, et al. (2014). High expression of 5-hydroxymethylcytosine and isocitrate dehydrogenase 2 is associated with favorable prognosis after curative resection of hepatocellular carcinoma. Journal of experimental & clinical cancer research : CR 33, 32. [PubMed: 24716838]

Lou Y, Diao L, Cuentas ER, Denning WL, Chen L, Fan YH, Byers LA, Wang J, Papadimitrakopoulou VA, Behrens C, et al. (2016). Epithelial-Mesenchymal Transition Is Associated with a Distinct Tumor Microenvironment Including Elevation of Inflammatory Signals and Multiple Immune Checkpoints in Lung Adenocarcinoma. Clinical cancer research : an official journal of the American Association for Cancer Research 22, 3630–3642. [PubMed: 26851185]

Louis DN, Perry A, Reifenberger G, von Deimling A, Figarella-Branger D, Cavenee WK, Ohgaki H, Wiestler OD, Kleihues P, and Ellison DW (2016). The 2016 World Health Organization Classification of Tumors of the Central Nervous System: a summary. Acta neuropathologica 131, 803–820. [PubMed: 27157931]

Lun AT, McCarthy DJ, and Marioni JC (2016). A step-by-step workflow for low-level analysis of single-cell RNA-seq data with Bioconductor. F1000Research 5, 2122. [PubMed: 27909575]

Ma QL, Wang JH, Wang YG, Hu C, Mu QT, Yu MX, Wang L, Wang DM, Yang M, Yin XF, et al. (2015). High IDH1 expression is associated with a poor prognosis in cytogenetically normal acute myeloid leukemia. International journal of cancer 137, 1058–1065. [PubMed: 25523507]

Ma W, Kim S, Chowdhury S, Li Z, Yang M, Yoo S, Petralia F, Jacobsen J, Li JJ, Ge X, et al. (2020). DreamAI: algorithm for the imputation of proteomics data. bioRxiv, 2020.2007.2021.214205.

MacDonald TJ, Aguilera D, and Kramm CM (2011). Treatment of high-grade glioma in children and adolescents. Neuro-oncology 13, 1049–1058. [PubMed: 21784756]

Magupalli VG, Mochida S, Yan J, Jiang X, Westenbroek RE, Nairn AC, Scheuer T, and Catterall WA (2013). Ca2+-independent activation of Ca2+/calmodulin-dependent protein kinase II bound to the C-terminal domain of CaV2.1 calcium channels. The Journal of biological chemistry 288, 4637–4648. [PubMed: 23255606]

Malta TM, Sokolov A, Gentles AJ, Burzykowski T, Poisson L, Weinstein JN, Kaminska B, Huelsken J, Omberg L, Gevaert O, et al. (2018). Machine Learning Identifies Stemness Features Associated with Oncogenic Dedifferentiation. Cell 173, 338–354.e315. [PubMed: 29625051]

Marie SK, Oba-Shinjo SM, da Silva R, Gimenez M, Nunes Reis G, Tassan JP, Rosa JC, and Uno M (2016). Stathmin involvement in the maternal embryonic leucine zipper kinase pathway in glioblastoma. Proteome science 14, 6. [PubMed: 26973435]

Meel MH, Schaper SA, Kaspers GJL, and Hulleman E (2018). Signaling pathways and mesenchymal transition in pediatric high-grade glioma. Cellular and molecular life sciences : CMLS 75, 871–887. [PubMed: 29164272]

Merlo LM, Pepper JW, Reid BJ, and Maley CC (2006). Cancer as an evolutionary and ecological process. Nature reviews Cancer 6, 924–935. [PubMed: 17109012]

Mertins P, Mani DR, Ruggles KV, Gillette MA, Clauser KR, Wang P, Wang X, Qiao JW, Cao S, Petralia F, et al. (2016). Proteogenomics connects somatic mutations to signalling in breast cancer. Nature 534, 55–62. [PubMed: 27251275]

Mertins P, Yang F, Liu T, Mani DR, Petyuk VA, Gillette MA, Clauser KR, Qiao JW, Gritsenko MA, Moore RJ, et al. (2014). Ischemia in tumors induces early and sustained phosphorylation changes in stress kinase pathways but does not affect global protein levels. Molecular & cellular proteomics : MCP 13, 1690–1704. [PubMed: 24719451]

Moniaux N, Nemos C, Schmied BM, Chauhan SC, Deb S, Morikane K, Choudhury A, Vanlith M, Sutherlin M, Sikela JM, et al. (2006). The human homologue of the RNA polymerase II-associated factor 1 (hPaf1), localized on the 19q13 amplicon, is associated with tumorigenesis. Oncogene 25, 3247–3257. [PubMed: 16491129]

Monroe ME, Shaw JL, Daly DS, Adkins JN, and Smith RD (2008). MASIC: a software program for fast quantitation and flexible visualization of chromatographic profiles from detected LC-MS(/MS) features. Computational biology and chemistry 32, 215–217. [PubMed: 18440872]

Morrissy AS, Garzia L, Shih DJ, Zuyderduyn S, Huang X, Skowron P, Remke M, Cavalli FM, Ramaswamy V, Lindsay PE, et al. (2016). Divergent clonal selection dominates medulloblastoma at recurrence. Nature 529, 351–357. [PubMed: 26760213]

Mounir M, Lucchetta M, Silva TC, Olsen C, Bontempi G, Chen X, Noushmehr H, Colaprico A, and Papaleo E (2019). New functionalities in the TCGAbiolinks package for the study and integration of cancer data from GDC and GTEx. PLoS computational biology 15, e1006701. [PubMed: 30835723]

Naus CC, Aftab Q, and Sin WC (2016). Common mechanisms linking connexin43 to neural progenitor cell migration and glioma invasion. Seminars in cell & developmental biology 50, 59–66. [PubMed: 26706148]

Navarrete-Perea J, Yu Q, Gygi SP, and Paulo JA (2018). Streamlined Tandem Mass Tag (SL-TMT) Protocol: An Efficient Strategy for Quantitative (Phospho)proteome Profiling Using Tandem Mass Tag-Synchronous Precursor Selection-MS3. Journal of proteome research 17, 2226–2236. [PubMed: 29734811]

Nesvizhskii AI, Keller A, Kolker E, and Aebersold R (2003). A statistical model for identifying proteins by tandem mass spectrometry. Analytical chemistry 75, 4646–4658. [PubMed: 14632076]

Newman ME, and Girvan M (2004). Finding and evaluating community structure in networks. Physical review E, Statistical, nonlinear, and soft matter physics 69, 026113.

Ng EL, and Tang BL (2008). Rab GTPases and their roles in brain neurons and glia. Brain research reviews 58, 236–246. [PubMed: 18485483]

Nie W, Xu MD, Gan L, Huang H, Xiu Q, and Li B (2015). Overexpression of stathmin 1 is a poor prognostic biomarker in non-small cell lung cancer. Laboratory investigation; a journal of technical methods and pathology 95, 56–64. [PubMed: 25384122]

Northcott PA, Buchhalter I, Morrissy AS, Hovestadt V, Weischenfeldt J, Ehrenberger T, Grobner S, Segura-Wang M, Zichner T, Rudneva VA, et al. (2017). The whole-genome landscape of medulloblastoma subtypes. Nature 547, 311–317. [PubMed: 28726821]

Ostrom QT, Gittleman H, Truitt G, Boscia A, Kruchko C, and Barnholtz-Sloan JS (2018). CBTRUS Statistical Report: Primary Brain and Other Central Nervous System Tumors Diagnosed in the United States in 2011–2015. Neuro-oncology 20, iv1–iv86. [PubMed: 30445539]

Panisko EA, and McAlister-Henn L (2001). Subunit interactions of yeast NAD+-specific isocitrate dehydrogenase. The Journal of biological chemistry 276, 1204–1210. [PubMed: 11042198]

Parsons DW, Li M, Zhang X, Jones S, Leary RJ, Lin JC, Boca SM, Carter H, Samayoa J, Bettegowda C, et al. (2011). The genetic landscape of the childhood cancer medulloblastoma. Science (New York, NY) 331, 435–439.

Pereira MSL, Klamt F, Thome CC, Worm PV, and de Oliveira DL (2017). Metabotropic glutamate receptors as a new therapeutic target for malignant gliomas. Oncotarget 8, 22279–22298. [PubMed: 28212543]

Perrot I, Michaud HA, Giraudon-Paoli M, Augier S, Docquier A, Gros L, Courtois R, Dejou C, Jecko D, Becquart O, et al. (2019). Blocking Antibodies Targeting the CD39/CD73 Immunosuppressive Pathway Unleash Immune Responses in Combination Cancer Therapies. Cell reports 27, 2411–2425.e2419. [PubMed: 31116985]

Petralia F, Song WM, Tu Z, and Wang P (2016). New Method for Joint Network Analysis Reveals Common and Different Coexpression Patterns among Genes and Proteins in Breast Cancer. Journal of proteome research 15, 743–754. [PubMed: 26733076]

Petralia F, Wang L, Peng J, Yan A, Zhu J, and Wang P (2018). A new method for constructing tumor specific gene co-expression networks based on samples with tumor purity heterogeneity. Bioinformatics (Oxford, England) 34, i528–i536.

Petralia F, Wang P, Yang J, and Tu Z (2015). Integrative random forest for gene regulatory network inference. Bioinformatics (Oxford, England) 31, i197–205.

Pilarczyk M, Najafabadi MF, 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.

Porro F, Rosato-Siri M, Leone E, Costessi L, Iaconcig A, Tongiorgi E, and Muro AF (2010). beta-adducin (Add2) KO mice show synaptic plasticity, motor coordination and behavioral deficits accompanied by changes in the expression and phosphorylation levels of the alpha- and gamma-adducin subunits. Genes, brain, and behavior 9, 84–96.

Pratilas CA, Taylor BS, Ye Q, Viale A, Sander C, Solit DB, and Rosen N (2009). (V600E)BRAF is associated with disabled feedback inhibition of RAF-MEK signaling and elevated transcriptional output of the pathway. Proceedings of the National Academy of Sciences of the United States of America 106, 4519–4524. [PubMed: 19251651]

Pugh TJ, Weeraratne SD, Archer TC, Pomeranz Krummel DA, Auclair D, Bochicchio J, Carneiro MO, Carter SL, Cibulskis K, Erlich RL, et al. (2012). Medulloblastoma exome sequencing uncovers subtype-specific somatic mutations. Nature 488, 106–110. [PubMed: 22820256]

Qiao S, Peng R, Yan H, Gao Y, Wang C, Wang S, Zou Y, Xu X, Zhao L, Dong J, et al. (2014). Reduction of phosphorylated synapsin I (ser-553) leads to spatial memory impairment by attenuating GABA release after microwave exposure in Wistar rats. PloS one 9, e95503. [PubMed: 24743689]

Rivero-Hinojosa S, Lau LS, Stampar M, Staal J, Zhang H, Gordish-Dressman H, Northcott PA, Pfister SM, Taylor MD, Brown KJ, et al. (2018). Proteomic analysis of Medulloblastoma reveals functional biology with translational potential. Acta neuropathologica communications 6, 48. [PubMed: 29880060]

Robbins D, Wittwer JA, Codarin S, Circu ML, Aw TY, Huang TT, Van Remmen H, Richardson A, Wang DB, Witt SN, et al. (2012). Isocitrate dehydrogenase 1 is downregulated during early skin

tumorigenesis which can be inhibited by overexpression of manganese superoxide dismutase. Cancer science 103, 1429–1433. [PubMed: 22533343]

Robinson G, Parker M, Kranenburg TA, Lu C, Chen X, Ding L, Phoenix TN, Hedlund E, Wei L, Zhu X, et al. (2012). Novel mutations target distinct subgroups of medulloblastoma. Nature 488, 43–48. [PubMed: 22722829]

Salomonis N, Dexheimer PJ, Omberg L, Schroll R, Bush S, Huo J, Schriml L, Ho Sui S, Keddache M, Mayhew C, et al. (2016). Integrated Genomic Analysis of Diverse Induced Pluripotent Stem Cells from the Progenitor Cell Biology Consortium. Stem cell reports 7, 110–125. [PubMed: 27293150]

Santamaria D, Barriere C, Cerqueira A, Hunt S, Tardy C, Newton K, Caceres JF, Dubus P, Malumbres M, and Barbacid M (2007). Cdk1 is sufficient to drive the mammalian cell cycle. Nature 448, 811–815. [PubMed: 17700700]

Savage SR, Shi Z, Liao Y, and Zhang B (2019). Graph Algorithms for Condensing and Consolidating Gene Set Analysis Results. Molecular & cellular proteomics : MCP 18, S141–s152. [PubMed: 31142576]

Schmitt JM, and Stork PJ (2001). Cyclic AMP-mediated inhibition of cell growth requires the small G protein Rap1. Molecular and cellular biology 21, 3671–3683. [PubMed: 11340161]

Schreck KC, Grossman SA, and Pratilas CA (2019). BRAF Mutations and the Utility of RAF and MEK Inhibitors in Primary Brain Tumors. Cancers 11.

Shah K, and Lahiri DK (2017). A Tale of the Good and Bad: Remodeling of the Microtubule Network in the Brain by Cdk5. Molecular neurobiology 54, 2255–2268. [PubMed: 26944284]

Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, and Ideker T (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome research 13, 2498–2504. [PubMed: 14597658]

Shin HJ, Lee S, and Jung HJ (2019). A curcumin derivative hydrazinobenzoylcurcumin suppresses stem-like features of glioblastoma cells by targeting Ca(2+) /calmodulin-dependent protein kinase II. Journal of cellular biochemistry 120, 6741–6752. [PubMed: 30390339]

Shteynberg DD, Deutsch EW, Campbell DS, Hoopmann MR, Kusebauch U, Lee D, Mendoza L, Midha MK, Sun Z, Whetton AD, et al. (2019). PTMProphet: Fast and Accurate Mass Modification Localization for the Trans-Proteomic Pipeline. Journal of proteome research 18, 4262–4272. [PubMed: 31290668]

Sokolov A, Paull EO, and Stuart JM (2016). ONE-CLASS DETECTION OF CELL STATES IN TUMOR SUBTYPES. Pacific Symposium on Biocomputing Pacific Symposium on Biocomputing 21, 405–416. [PubMed: 26776204]

Solga AC, Pong WW, Walker J, Wylie T, Magrini V, Apicelli AJ, Griffith M, Griffith OL, Kohsaka S, Wu GF, et al. (2015). RNA-sequencing reveals oligodendrocyte and neuronal transcripts in microglia relevant to central nervous system disease. Glia 63, 531–548. [PubMed: 25258010]

Song X, Ji J, Gleason KJ, Yang F, Martignetti JA, Chen LS, 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. Molecular & cellular proteomics : MCP 18, S52–s65. [PubMed: 31227599]

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 research 48, D431–D439.

Stenson PD, Mort M, Ball EV, Evans K, Hayden M, Heywood S, Hussain M, Phillips AD, and Cooper DN (2017). The Human Gene Mutation Database: towards a comprehensive repository of inherited mutation data for medical research, genetic diagnosis and next-generation sequencing studies. Human genetics 136, 665–677. [PubMed: 28349240]

Stepulak A, Rola R, Polberg K, and Ikonomidou C (2014). Glutamate and its receptors in cancer. Journal of neural transmission (Vienna, Austria : 1996) 121, 933–944.

Subramanian A, Narayan R, Corsello SM, Peck DD, Natoli TE, Lu X, Gould J, Davis JF, Tubelli AA, Asiedu JK, et al. (2017). A Next Generation Connectivity Map: L1000 Platform and the First 1,000,000 Profiles. Cell 171, 1437–1452.e1417. [PubMed: 29195078]

Sun Q, Guo S, Wang CC, Sun X, Wang D, Xu N, Jin SF, and Li KZ (2015). Cross-talk between TGF-beta/Smad pathway and Wnt/beta-catenin pathway in pathological scar formation. International journal of clinical and experimental pathology 8, 7631–7639. [PubMed: 26261683]

Talevich E, Shain AH, Botton T, and Bastian BC (2016). CNVkit: Genome-Wide Copy Number Detection and Visualization from Targeted DNA Sequencing. PLoS computational biology 12, e1004873. [PubMed: 27100738]

Tanaka H, Sasayama T, Tanaka K, Nakamizo S, Nishihara M, Mizukawa K, Kohta M, Koyama J, Miyake S, Taniguchi M, et al. (2013). MicroRNA-183 upregulates HIF-1α by targeting isocitrate dehydrogenase 2 (IDH2) in glioma cells. Journal of neuro-oncology 111, 273–283. [PubMed: 23263745]

Tian GY, Zang SF, Wang L, Luo Y, Shi JP, and Lou GQ (2015). Isocitrate Dehydrogenase 2 Suppresses the Invasion of Hepatocellular Carcinoma Cells via Matrix Metalloproteinase 9. Cellular physiology and biochemistry : international journal of experimental cellular physiology, biochemistry, and pharmacology 37, 2405–2414.

Tomson BN, and Arndt KM (2013). The many roles of the conserved eukaryotic Paf1 complex in regulating transcription, histone modifications, and disease states. Biochimica et biophysica acta 1829, 116–126. [PubMed: 22982193]

Turnescu T, Arter J, Reiprich S, Tamm ER, Waisman A, and Wegner M (2018). Sox8 and Sox10 jointly maintain myelin gene expression in oligodendrocytes. Glia 66, 279–294. [PubMed: 29023979]

Tusher VG, Tibshirani R, and Chu G (2001). Significance analysis of microarrays applied to the ionizing radiation response. Proceedings of the National Academy of Sciences of the United States of America 98, 5116–5121. [PubMed: 11309499]

Uchida S, Martel G, Pavlowsky A, Takizawa S, Hevi C, Watanabe Y, Kandel ER, Alarcon JM, and Shumyatsky GP (2014). Learning-induced and stathmin-dependent changes in microtubule stability are critical for memory and disrupted in ageing. Nature communications 5, 4389.

Venkataramani V, Tanev DI, Strahle C, Studier-Fischer A, Fankhauser L, Kessler T, Korber C, Kardorff M, Ratliff M, Xie R, et al. (2019). Glutamatergic synaptic input to glioma cells drives brain tumour progression. Nature 573, 532–538. [PubMed: 31534219]

Venkatesh HS, Morishita W, Geraghty AC, Silverbush D, Gillespie SM, Arzt M, Tam LT, Espenel C, Ponnuswami A, Ni L, et al. (2019). Electrical and synaptic integration of glioma into neural circuits. Nature 573, 539–545. [PubMed: 31534222]

Vu NT, Park MA, Shultz JC, Goehe RW, Hoeferlin LA, Shultz MD, Smith SA, Lynch KW, and Chalfant CE (2013). hnRNP U enhances caspase-9 splicing and is modulated by AKT-dependent phosphorylation of hnRNP L. The Journal of biological chemistry 288, 8575–8584. [PubMed: 23396972]

Wang L, Karpova A, Gritsenko MA, Kyle JE, Cao S, Rykunov D, Li Y, Colaprico A, Rothstein J, Hong R, et al. (2020). Proteogenomic and Metabolomic Characterization of Human Glioblastoma. Cell under review.

Wang X, Park J, Susztak K, Zhang NR, and Li M (2019). Bulk tissue cell type deconvolution with multi-subject single-cell expression reference. Nature communications 10, 380.

Whiteaker JR, Zhao L, Saul R, Kaczmarczyk JA, Schoenherr RM, Moore HD, Jones-Weinert C, Ivey RG, Lin C, Hiltke T, et al. (2018). A Multiplexed Mass Spectrometry-Based Assay for Robust Quantification of Phosphosignaling in Response to DNA Damage. Radiation research 189, 505–518. [PubMed: 29474155]

Wilkerson MD, and Hayes DN (2010). ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. Bioinformatics (Oxford, England) 26, 1572–1573.

Wittstatt J, Reiprich S, and Kuspert M (2019). Crazy Little Thing Called Sox-New Insights in Oligodendroglial Sox Protein Function. International journal of molecular sciences 20.

Yoo S, Huang T, Campbell JD, Lee E, Tu Z, Geraci MW, Powell CA, Schadt EE, Spira A, and Zhu J (2014). MODMatcher: multi-omics data matcher for integrative genomic analysis. PLoS computational biology 10, e1003790. [PubMed: 25122495]

Yoshihara K, Shahmoradgoli M, Martinez E, Vegesna R, Kim H, Torres-Garcia W, Trevino V, Shen H, Laird PW, Levine DA, et al. (2013). Inferring tumour purity and stromal and immune cell admixture from expression data. Nature communications 4, 2612.

Yu K, Lin CJ, Hatcher A, Lozzi B, Kong K, Huang-Hobbs E, Cheng YT, Beechar VB, Zhu W, Zhang Y, et al. (2020). PIK3CA variants selectively initiate brain hyperactivity during gliomagenesis. Nature 578, 166–171. [PubMed: 31996845]

Yuan J, Levitin HM, Frattini V, Bush EC, Boyett DM, Samanamud J, Ceccarelli M, Dovas A, Zanazzi G, Canoll P, et al. (2018). Single-cell transcriptome analysis of lineage diversity in high-grade glioma. Genome medicine 10, 57. [PubMed: 30041684]

Zecha J, Satpathy S, Kanashova T, Avanessian SC, Kane MH, Clauser KR, Mertins P, Carr SA, and Kuster B (2019). TMT Labeling for the Masses: A Robust and Cost-efficient, In-solution Labeling Approach. Molecular & cellular proteomics : MCP 18, 1468–1478. [PubMed: 30967486]

Zhang B, Wang J, Wang X, Zhu J, Liu Q, Shi Z, Chambers MC, Zimmerman LJ, Shaddox KF, Kim S, et al. (2014). Proteogenomic characterization of human colon and rectal cancer. Nature 513, 382–387. [PubMed: 25043054]

Zhang H, Liu T, Zhang Z, Payne SH, Zhang B, McDermott JE, Zhou JY, Petyuk VA, Chen L, Ray D, et al. (2016). Integrated Proteogenomic Characterization of Human High-Grade Serous Ovarian Cancer. Cell 166, 755–765. [PubMed: 27372738]

Zhang Y, Lv W, Li Q, Wang Q, Ru Y, Xiong X, Yan F, Pan T, Lin W, and Li X (2019). IDH2 compensates for IDH1 mutation to maintain cell survival under hypoxic conditions in IDH1−mutant tumor cells. Molecular medicine reports 20, 1893–1900. [PubMed: 31257503]

Zhao L, Whiteaker JR, Pope ME, Kuhn E, Jackson A, Anderson NL, Pearson TW, Carr SA, and Paulovich AG (2011). Quantification of proteins using peptide immunoaffinity enrichment coupled with mass spectrometry. Journal of visualized experiments : JoVE.
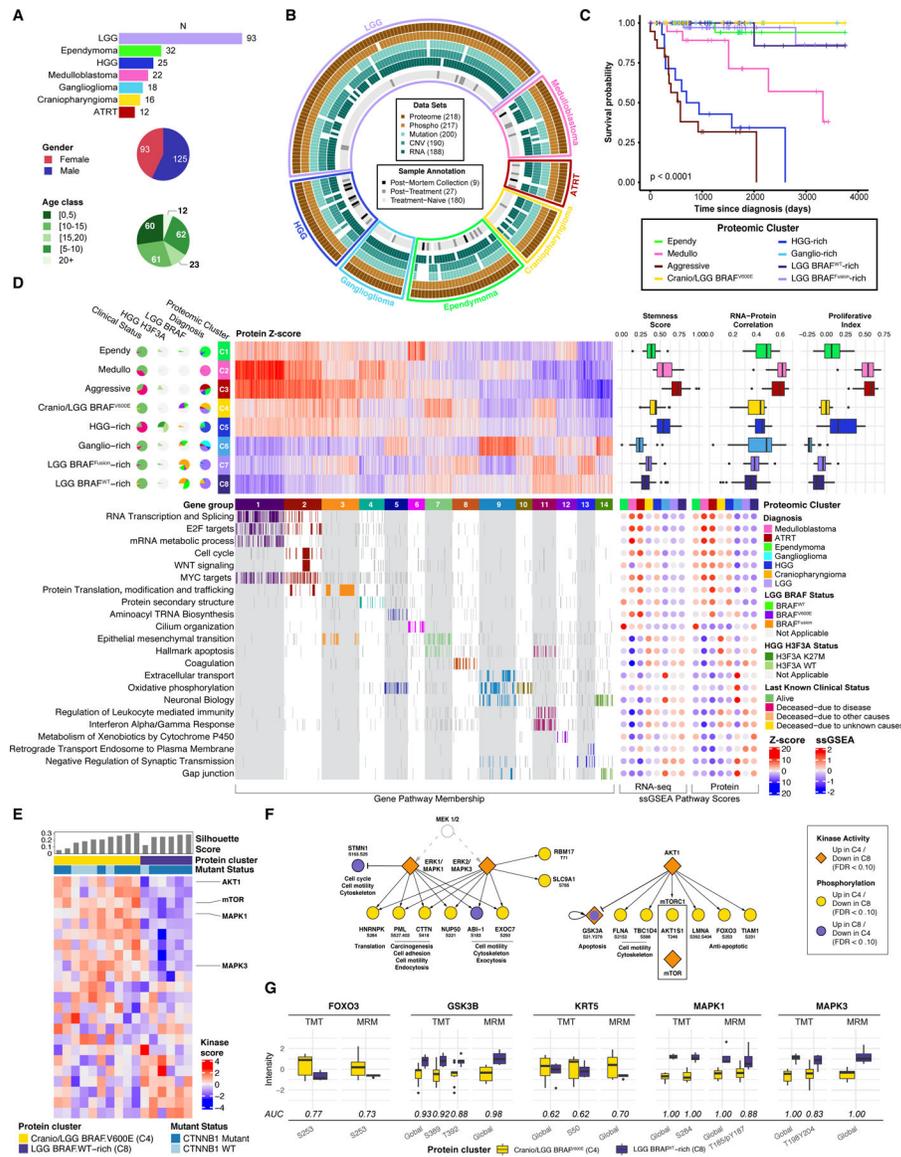
**Figure 1. Proteomic based clustering of pediatric brain tumors.**

**A.** Summary of pediatric brain tumor cohort.

**B.** Presence of omics data sets for each of the 218 tumor samples. For each sample, the clinical status at sample collection (i.e., post- mortem, post-treatment or treatment naive) is also reported.

**C.** Kaplan Meier curves for OS of patients stratified by proteomic cluster.

**D.** Proteomic clusters and differentially expressed proteins allocated to 14 gene clusters (top heatmap). Each row represents a proteomic cluster, while each column represents a protein. Red/blue colors denote up/down regulation patterns of different proteins in a cluster. Distributions of diagnoses, clinical outcomes, and mutation status among the 8 clusters (top left pie plots), and gene members of key pathways enriched in each gene group (bottom heatmap) are shown. For each pathway, the averaged ssGSEA score in each protein cluster

based on global proteomic (Protein) and RNA-seq (RNA-seq) data are illustrated to the right.

**E.** Heatmap of kinase activity scores for the CP tumors (n=16). Silhouette scores (top) measures the cohesiveness of tumors classified as C4 and C8 based on kinase activity score. Kinases involved in AKT1 or ERK1/2 signaling are highlighted in the heatmap.

**F.** Diagram illustrating differences between C4 and C8 CP tumors in terms of phosphorylation abundance and kinase activity for AKT and ERK1/2 signaling members.

**G.** MRM measurements validated different activities of proteins and phosphoproteins between C4 and C8 CPs. The numbers annotated under each pair of boxplots correspond to AUC (area under the curve) for classifying the two groups of CP using the corresponding protein/phosphosite measurement.
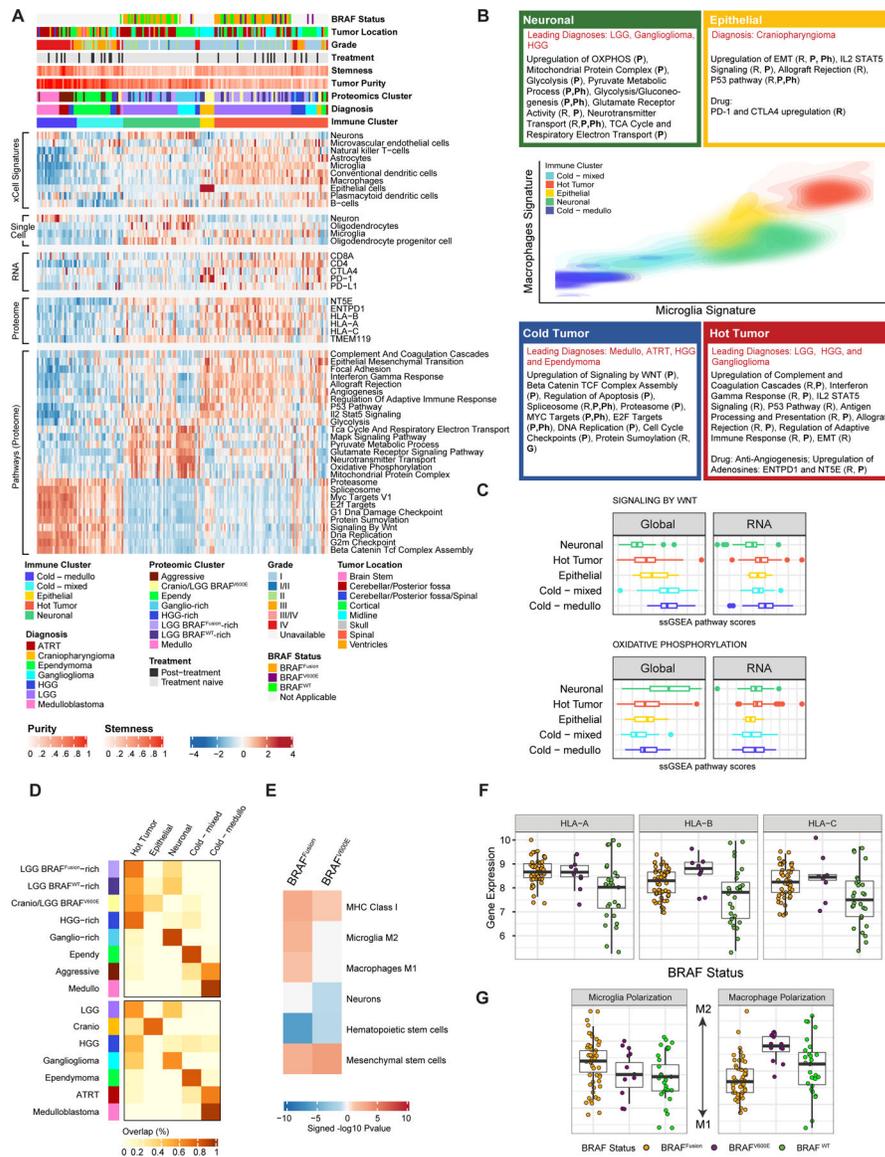
**Figure 2. Immune infiltration in pediatric brain tumors.**

**A.** Heatmap illustrating cell type compositions, and activities of selected individual gene/proteins and pathways across 5 immune clusters. The heatmap in the first section illustrates the immune/stromal signatures from xCell. The heatmap in the second section illustrates signatures of microglia, neurons and oligodendrocytes derived from single cell sequencing data from Darmanis et al (2017). RNA and protein abundance of key immune-related markers, and ssGSEA scores based on global proteomic data for biological pathways upregulated in different immune groups are illustrated in the remaining sections.

**B.** Contour plot of two-dimensional density based on Macrophage (y-axis) and Microglia scores (x-axis) for different immune clusters. For each immune cluster, key upregulated pathways significant at 10% FDR are reported based on RNAseq (R), global proteomic (P) and phospho-proteomic data (Ph) in the annotation boxes. For Cold-mixed and Cold-medullo clusters, pathways upregulated in both clusters are reported.

**C.** Distribution of pathway scores of Signaling by WNT and Oxidative Phosphorylation based on global proteomic data and RNA stratified by immune clusters.

**D.** Heatmap showing the comparison between immune clusters (columns) with proteomic clusters and different histologies (rows). Each row sums to one, with different entries showing the proportion of tumors allocated to different immune clusters.

**E.** xCell immune/stromal and antigen presentation signatures in $BRAF^{V600E}$ or $BRAF^{Fusion}$ compared to $BRAF^{WT}$ in LGG.

**F.** Distribution of RNA levels of $HLA\text{-}A$, $HLA\text{-}B$ and $HLA\text{-}C$ in LGG tumors with different $BRAF$ statuses.

**G.** Distribution of macrophage and microglia polarization (M2-M1) in LGG tumors with different $BRAF$ statuses.
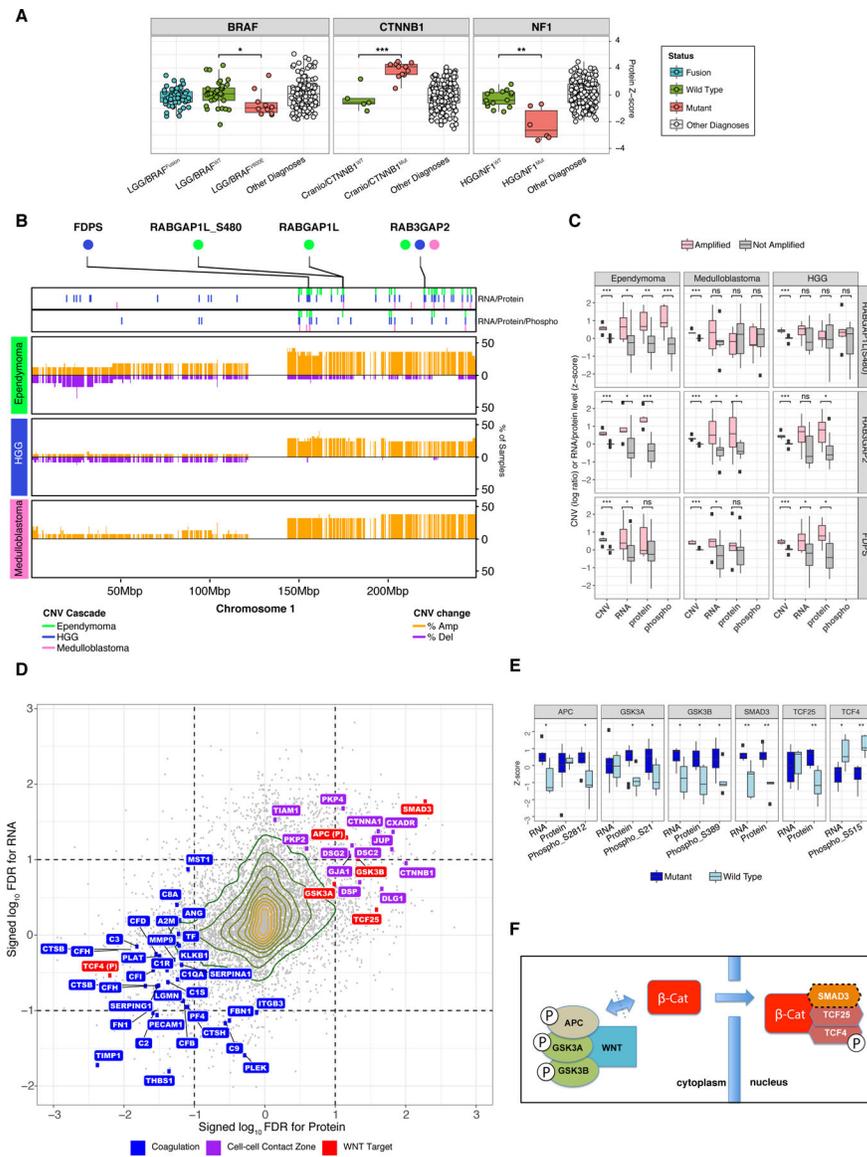
**Figure 3. Impact of genomic alterations on transcriptomic, proteomic and phosphoproteomic abundances.**

**A.** Distribution of protein abundance of BRAF, CTNNB1, and NF1 across tumor samples stratified by different mutation status and diagnoses. Symbols *, **, and *** correspond to p-values less than 0.1, 0.01 and 0.001, respectively.

**B.** DNA copy number amplification/deletion frequencies along chromosome 1 among EP, HGG and MB samples. Genes with detected CNV-RNA/protein or CNV-RNA/protein/ phospho cascade events are labelled as vertical bars in the top track.

**C.** Distribution of DNA copy number (log ratio), RNA and protein abundance of RABGAP1L, RAB3GAP2 and FDPS stratified by their amplification statuses in EP, MB and HGG tumors. For RABGAP1L, Symbols *, ** and *** mean the same as in A. "ns" stands for "not significant" (p-value>0.1).

**D.** Illustration of the impact of *CTNNB1* mutation on RNA and protein abundance in CP samples. x-axis (y-axis) represents signed -log10 FDR for testing the association between

protein abundances (RNAs) and *CTNNB1* mutation. Cell-Cell Contact Zone (Coagulation) pathway is enriched in the set of proteins up (down) regulated in *CTNNB1* mutant samples. A few members of the WNT Signaling pathway whose protein or phosphosites are associated with *CTNNB1* mutation are highlighted in red. Phosphosites are annotated with "(P)" in their gene symbols.

**E.** Distribution of protein and phosphosite abundances among *CTNNB1* mutant and *CTNNB1* wild-type CP tumors for known key members of the WNT Signaling pathway interacting with β-Catenin and transcription factors regulated by *CTNNB1*. Symbols *, ** and *** correspond to FDR less than 0.1, 0.01 and 0.001, respectively. "ns" stands for "not significant" (FDR >0.1).

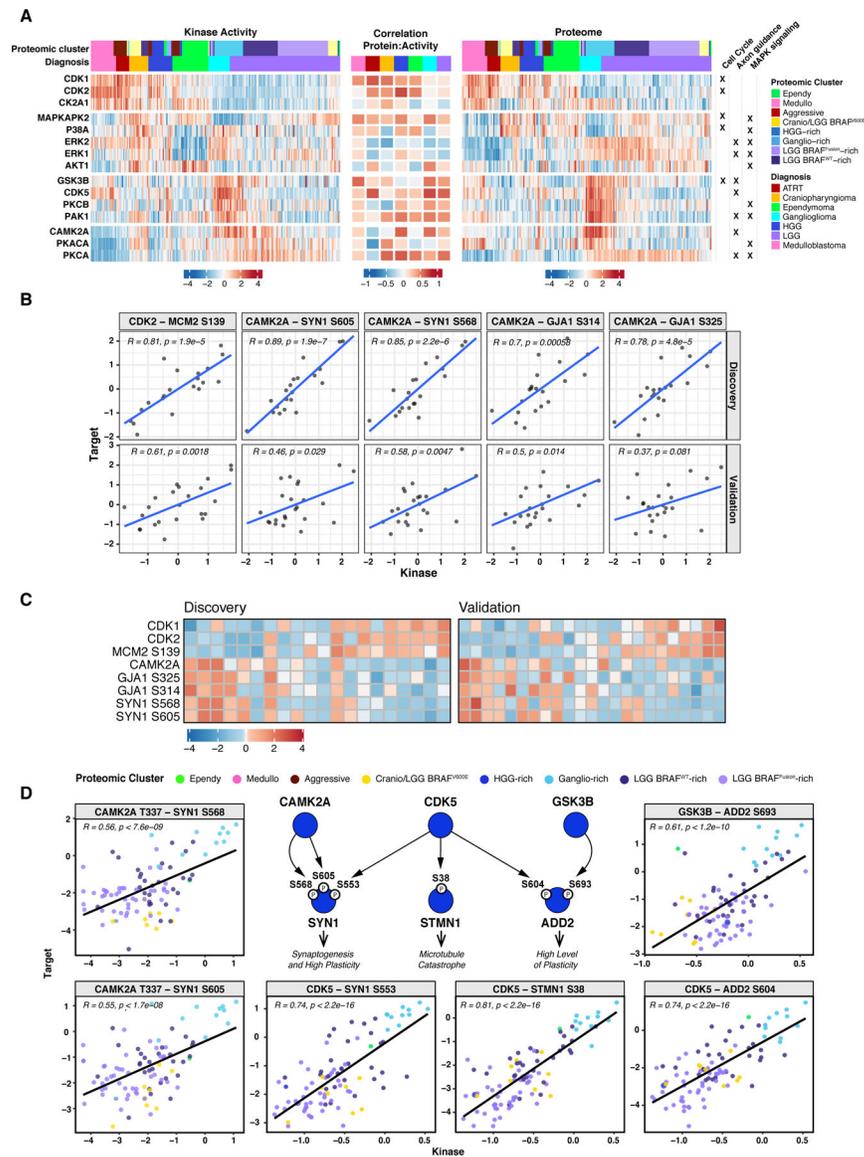**F.** Illustration of the regulatory role of β-Catenin.

**Figure 4. Phosphoproteomic analysis of kinase activity**

**A.** Heatmaps showing the global abundance (right panel) and the kinase activity score (left panel) of selected kinases across different histologies. For each kinase, the Pearson's correlation between its global abundance and kinase activity within each histology is shown in the middle panel.

**B.** Scatterplot showing the global abundance of a particular kinase (x-axis) versus the phospho-abundance of the targeted substrates (y-axis). First row is based on the data from the discovery cohort; while the second row displays the data based on the validation cohort.

**C.** Heatmap showing global proteomic abundance of CDK1, CDK2 and CAMK2A as well as phosphorylation abundance of MCM2 Ser 139, GJA1 Ser 325, GJA1 Ser 314, SYN1 Ser 568 and SYN1 Ser 605 among HGG in the discovery and validation cohorts.

**D.** Diagram showing kinase-substrate associations involved in CNS development in LGG (top-middle panel). Scatter plots showing the association between the global (or phospho)

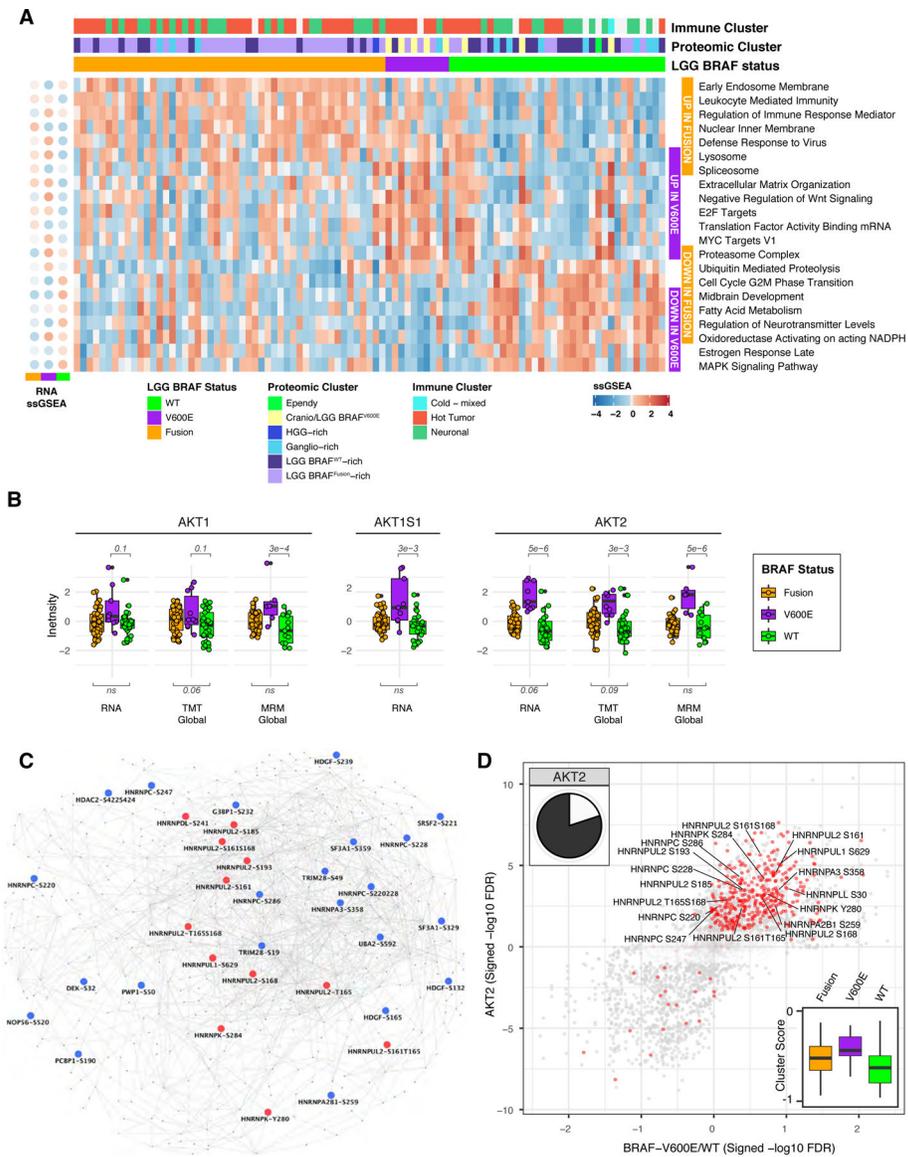abundance of each kinase (x-axis) and the phospho-abundance of the corresponding substrate (y-axis).

**Figure 5. Insights from proteogenomic analysis of LGG**

**A.** Heatmap illustrating ssGSEA scores of selected pathways differentially expressed between LGG tumors with different *BRAF* statuses based on global proteomic data. Dot-plot on the left-side summarizes ssGSEA pathway scores based on RNA data among samples with different *BRAF* statuses.

**B.** Distributions of RNA, TMT protein abundance (TMT Global), and MRM protein abundance (MRM Global) of AKT1, AKT2, and AKT1S1 in samples with different *BRAF* alteration statuses. FDR levels of two-sample comparisons between $BRAF^{V600E}/$
$BRAF^{Fusion}$ and $BRAF^{WT}$ are annotated.

**C.** The network topology representing the LGG phosphosite co-expression network module enriched in sites upregulated in $BRAF^{v600E}$ compared to $BRAF^{WT}$ tumors. Phosphosites mapping to genes in the HNRNP family or contained in the MYC Targets pathway are highlighted in red and blue, respectively.

**D.** Scatterplot displaying the association between each phosphosite's abundance with the global abundance of AKT2 (y-axis) versus the association with $BRAF^{V600E}$ (x-axis). Phosphosites contained in the network module in **C** are highlighted in red. Boxplots illustrate the distribution of the activity scores (ssGSEA) of the network module in **C** based on phosphoproteomic data in samples with different BRAF status. Pie-plot shows the proportion of phosphosites contained in the network module in C whose abundances are associated at 5% FDR with the global abundance of AKT2.

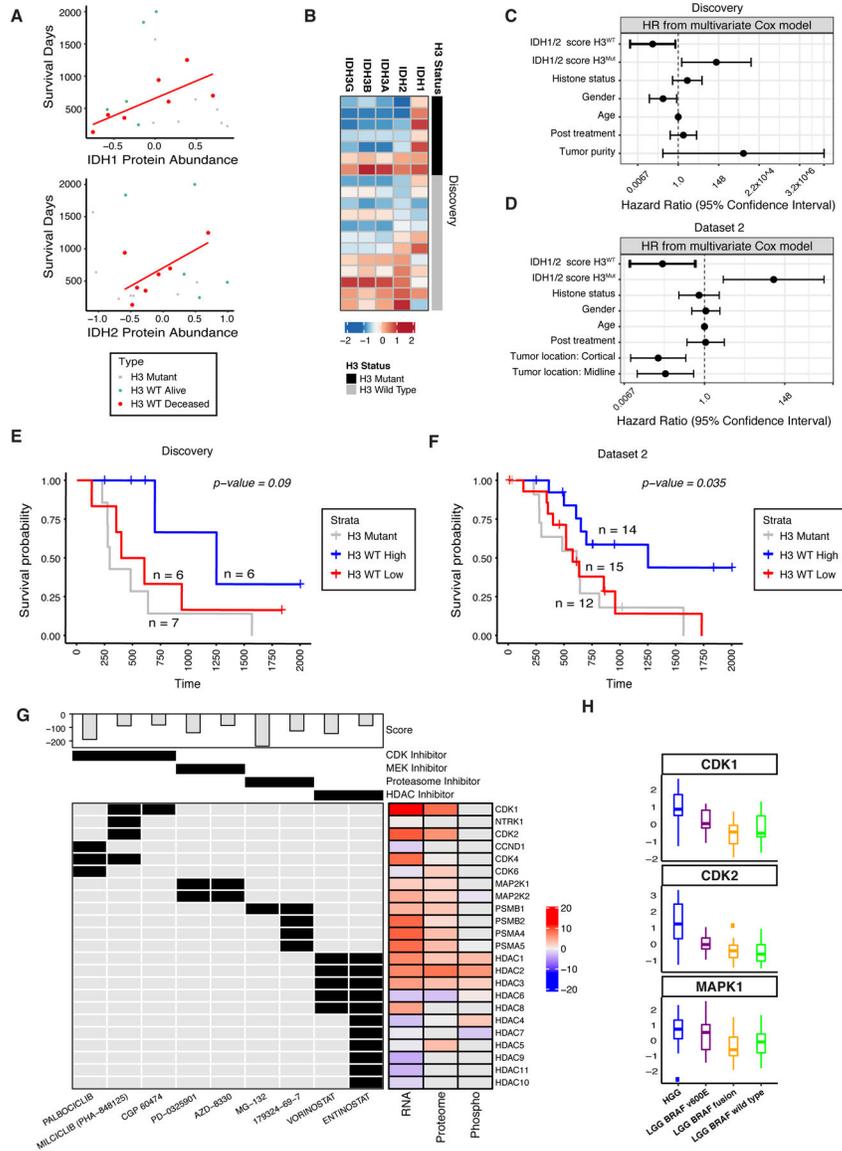**Figure 6. Insights from proteogenomic analysis of HGG**

**A.** Scatterplot showing OS of HGG patients versus the global protein abundance of IDH1 and IDH2 in the tumors.

**B.** Heatmap of global abundance of IDH proteins in the discovery cohort.

**C, D.** 95% CI of hazard ratio coefficients from Cox-regression for IDH1/2 scores and other covariates based on the discovery cohort (panel C) and Data Set 2 (D).

**E, F.** Kaplan-Meier curves of overall survival for HGG $H3^{Mut}$ samples (grey), $H3^{WT}$ samples with low IDH1/2 abundance (red) and $H3^{WT}$ tumors with high IDH1/2 abundance (blue) for the discovery cohort (panel **E**) and the validation cohort (**F**).**G.** Illustration of drug target analysis result. The bottom-left heatmap illustrates the targeting genes (rows) of each detected drugs (columns). For each gene, the z-score comparing its RNA and proteomic abundances between HGG and LGG is shown in the bottom-right heatmap. Mechanism of

actions are annotated on the top of the heatmap together with the resulting score from the cMAP analysis.

**H.** Distribution of kinase activity scores of CDK1, CDK2 and MAPK1 among HGG and LGG tumors, with the latter further stratified by BRAF status.
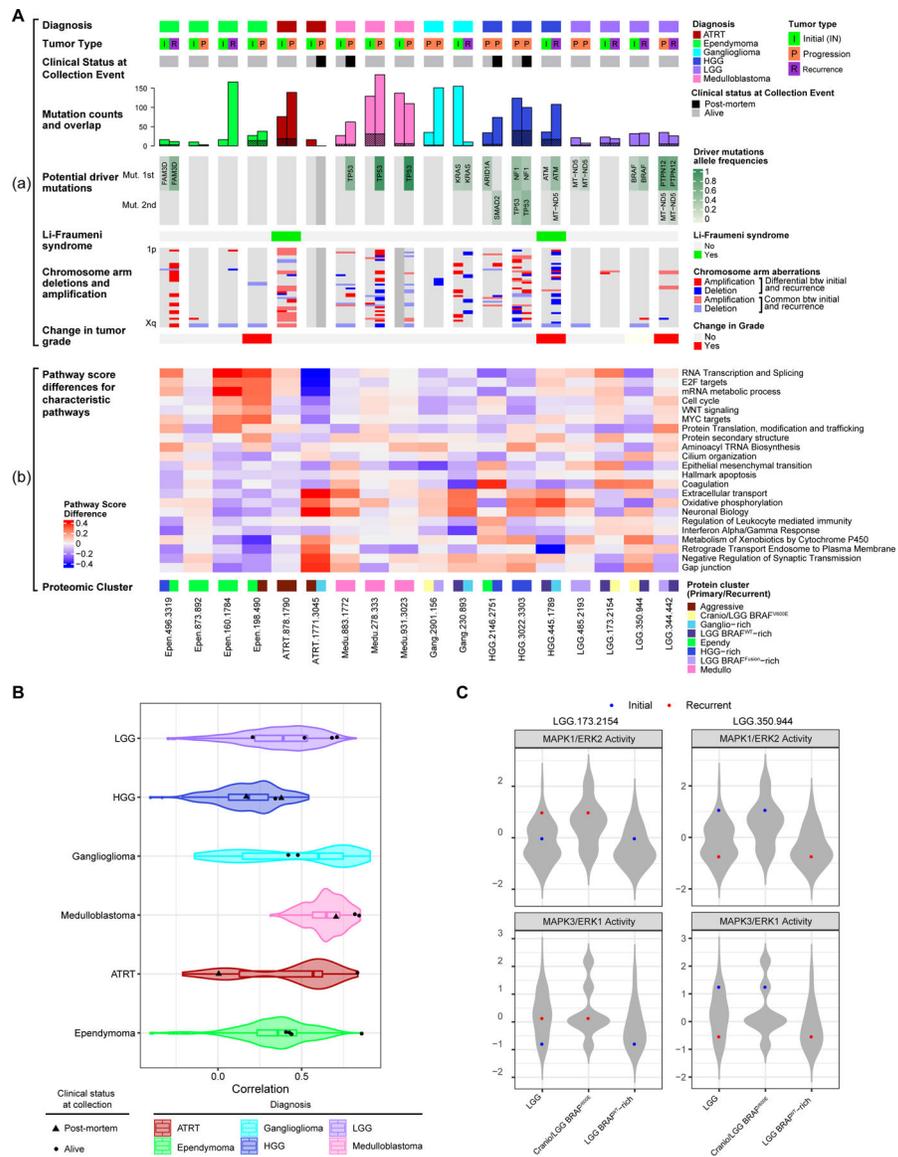
**Figure 7. Comparison between Initial and Recurrent Tumors**

**A. (a)** Clinical properties and genomic characterization of 18 pairs of IN vs RP tumors. The bar plot illustrates the number of non-synonymous mutations in IN and RP tumors with the number of shared mutations being represented by the shaded area. The potential driver mutation track shows the allele frequencies of somatic mutations of known oncogenes and tumor suppressor genes. Chromosome arm aberrations of each sample and the change of tumor grade from IN to RP of each patient are also shown. **(b)** Differences in ssGSEA score between RP and IN tumors of key molecular pathways associated with different proteomic clusters. The annotation at the bottom indicates the diagnosis and clinical event IDs of the paired samples for each patient. For example, "Epen.496.3319" refers to a pair of EP tumors with IDs: 7316–496 and 7316–3319.

**B.** Distribution of Spearman's correlation between the proteomic abundance of any pair of tumors within a particular histology. Correlations between the 18 paired IN-RP samples were further labeled in the violin plots.

**C.** Distribution of kinase activity scores of MAPK1/3 among all LGG samples, LGG samples allocated to C4 and LGG samples allocated to C8. IN and RP samples of patients LGG.350.944 and LGG.173.2154 are highlighted.