UC San Diego

Title

Proteomics, Post-translational Modifications, and Integrative Analyses Reveal Molecular Heterogeneity within Medulloblastoma Subgroups

Permalink

https://escholarship.org/uc/item/2hb1w351

Journal

ISSN

Authors

Archer, Tenley C
Ehrenberger, Tobias
Mundt, Filip
et al.

Publication Date

DOI

Proteomics, Post-translational Modifications, and Integrative Analyses Reveal Molecular Heterogeneity within Medulloblastoma Subgroups

Tenley C. Archer[#][1,2,3,4], Tobias Ehrenberger[#][5], Filip Mundt[#][4,6], Maxwell P. Gold[5], Karsten Krug[4], Clarence K. Mah[7], Elizabeth L. Mahoney[1,2], Colin J. Daniel[8], Alexander LeNail[5], Divya Ramamoorthy[5], Philipp Mertins[4,9,10], D. R. Mani[4], Hailei Zhang[4], Michael A. Gillette[2,4,11], Karl Clauser[4], Michael Noble[4], Lauren C. Tang[4], Jessica Pierre-François[1,2], Jacob Silterra[4], James Jensen[7], Pablo Tamayo[7,12], Andrey Korshunov[13,14], Stefan M. Pfister[15,16,17], Marcel Kool[15,16], Paul A. Northcott[16,18], Rosalie C. Sears[8], Jonathan O. Lipton[1,2,3], Steven A. Carr[4,§], Jill P. Mesirov[7,12,§], Scott L. Pomeroy[1,2,4,§], and Ernest Fraenkel[4,5,§,*]

[1]Department of Neurology, Boston Children's Hospital, Boston, MA, USA.

[2]Harvard Medical School, Boston, MA, USA.

[3]Division of Sleep Medicine, Harvard Medical School, Boston, MA, USA.

[4]Eli and Edythe Broad Institute of MIT and Harvard, Cambridge, MA, USA.

[5]Department of Biological Engineering, Massachusetts Institute of Technology (MIT), Cambridge, MA, USA.

[6]Department of Oncology-Pathology, Karolinska Institute, Stockholm, Sweden.

[7]Department of Medicine, University of California San Diego (UCSD), La Jolla, CA, USA.

[8]Department of Molecular and Medical Genetics, Oregon Health and Science University (OHSU), Portland, OR, USA.

[9]Present address: Proteomics Platform, Max Delbrück Center for Molecular Medicine in the Helmholtz Society, Berlin, Germany.

[10]Present address: Berlin Institute of Health, Berlin, Germany.

[11]Division of Pulmonary and Critical Care Medicine, Massachusetts General Hospital (MGH), Boston, MA, USA.

Declaration of Interests:

The authors declare no competing interests.

[12]Moores Cancer Center, University of California San Diego (UCSD), La Jolla, CA, USA.

[13]CCU Neuropathology, German Cancer Research Center (DKFZ), Heidelberg, Germany.

[14]Department of Neuropathology, Heidelberg University, Heidelberg, Germany.

[15]Hopp Children's Cancer Center at the NCT Heidelberg (KiTZ), Heidelberg, Germany.

[16]Division of Pediatric Neurooncology, German Cancer Research Center (DKFZ) and German Cancer Consortium (DKTK), Heidelberg, Germany.

[17]Department of Hematology and Oncology, Heidelberg University Hospital, Heidelberg, Germany.

[18]Department of Developmental Neurobiology, St Jude Children's Research Hospital, Memphis, TN, USA.

[#] These authors contributed equally to this work.

Summary

There is a pressing need to identify therapeutic targets in tumors with low mutation rates such as the malignant pediatric brain tumor medulloblastoma. To address this challenge, we quantitatively profiled global proteomes and phospho-proteomes of 45 medulloblastoma samples. Integrated analyses revealed that tumors with similar RNA expression vary extensively at the post-transcriptional and post-translational levels. We identified distinct pathways associated with two subsets of SHH tumors, and found post-translational modifications of MYC that are associated with poor outcomes in Group 3 tumors. We found kinases associated with subtypes and showed that inhibiting PRKDC sensitizes MYC-driven cells to radiation. Our study shows that proteomics enables a more comprehensive, functional readout, providing a foundation for future therapeutic strategies.

Significance

Genomic and epigenomic analyses have revolutionized cancer diagnostics. Nevertheless, it has been difficult to identify therapeutic targets for tumors that lack recurrent genomic lesions. Here we used global, mass spectrometry-based measurements of protein levels and post-translational modifications to identify functional pathways associated with subtypes of medulloblastoma. Strong proteomic signals revealed altered pathways that were not detected transcriptionally. One molecular subgroup of tumors showed marked discordance of RNA and protein levels, suggesting global changes in translation and/or proteostasis. We demonstrate the utility of an integrative approach for discovery of candidate biomarkers or drug targets and provide a multi-omic dataset that will serve as a resource for the community. This study has the potential to impact clinical trial design.

Introduction

Medulloblastoma is one of the most common pediatric brain tumors. Survival rates are high, but current therapies can leave lasting side effects including problems with speech, cognition and behavior, and increased risks of secondary cancers (Archer et al., 2017). Identifying the

pathways driving medulloblastoma could guide development of less toxic and more effective targeted therapies. Previous analyses have demonstrated that, at the molecular level, medulloblastoma is extremely heterogeneous and comprises at least four major consensus subgroups: wingless (WNT), sonic hedgehog (SHH), Group 3, and Group 4 (Cho et al., 2011; Hovestadt et al., 2014; Kool et al., 2012; Pugh et al., 2012; Robinson et al., 2012; Tamayo et al., 2011; Taylor et al., 2012). Almost all WNT tumors carry activating mutations in the β-catenin gene (*CTNNB1*) (Pugh et al., 2012). Unfortunately, the ubiquity of WNT signaling in non-cancer cells makes this pathway a challenging one for targeted cancer therapeutics (Kahn, 2014). Some SHH tumors respond to SMO inhibitors (Kool et al., 2014; Robinson et al., 2015). Group 3 and Group 4 are the least understood subgroups and constitute more than half of all medulloblastoma cases. These tumors have few consistent genetic abnormalities amenable to currently available targeted therapies (Northcott et al., 2012).

Deeper analysis of the transcriptome and epigenome has revealed subtypes within the four consensus subgroups as defined by Taylor *et al.* (2012). Based on transcriptome evidence, we reported two subtypes of Group 3 tumors, including one with a dominant MYC-driven signature and exceedingly poor prognosis (Cho et al., 2011; Pfister et al., 2009; Schwalbe et al., 2017). More recently, Northcott *et al.* (2017) defined eight subtypes of Group 3 and Group 4 tumors based on DNA methylation, including a category of MYC-driven samples called subtype II. Similarly, Cavalli *et al.* (2017) proposed a division of Group 3 and Group 4 into a total of six different subtypes. Recent evidence suggests that the MYCdriven Group 3 tumors may be susceptible to CDK inhibitors or to combinations of PI3K and HDAC inhibitors (Hanaford et al., 2016; Pei et al., 2016).

Recent advances in mass spectrometry have allowed for proteomic and phospho-proteomic profiling of other cancers, and integration of these data with other biological data developed a more complete understanding of specific cancers and their genetic drivers (Edwards et al., 2015; Huang et al., 2017; Lawrence et al., 2015; Mertins et al., 2016; Zhang et al., 2014, 2016). Since proteins are ultimately the functional effectors of biological activity in cancer cells, we hypothesized that global proteomic analysis may be an especially sensitive method for identifying potential therapeutic targets in medulloblastoma.

Results

Global proteomics reveals medulloblastoma subgroups

To obtain a comprehensive view of medulloblastoma, we selected 45 primary tumors from all consensus subgroups in a cohort that had been previously characterized (Cho et al., 2011; Hovestadt et al., 2014; Kool et al., 2012; Northcott et al., 2017). Of these, 42 had been analyzed for DNA methylation, 41 had preexisting whole genome sequencing and 39 had RNA-seq data (Figure 1, Tables S1 and S2) (Northcott et al., 2017).

Towards our aim of understanding the functional pathways that distinguish medulloblastoma subgroups, we collected proteomic, phospho-proteomic, and protein acetylation data for all 45 samples using isobaric labeling with TMT10 mass-tag reagents (Rauniyar and Yates 3rd, 2014) followed by high-resolution liquid chromatography-tandem mass spectrometry.

Phosphorylation events were enriched using metal-affinity enrichment (detecting phosphorylated serine, threonine, and tyrosine peptides: pSTY). In addition, we used antibodies to enrich for phosphorylated tyrosine peptides (pY) and acetylated lysine residues (acK). Over 13,000 proteins, more than 50,000 phosphosites, and almost 11,000 acetylated sites were quantified in total. Unless otherwise noted, only the complete data sets, in which proteomic features had been measured across all 45 samples, were used for analyses (Figure 1).

We applied uniform data normalization methods to each type of data and clustered samples using each data type separately (Figure 2A). Consensus clustering, principal component analysis (PCA), and t-distributed stochastic neighbor embedding (t-SNE) carried out separately on each data type predominantly revealed the known subgroups: Group 3, Group 4, and SHH (Figures 2A, S1-S3). The proteomic data also identified very stable subsets of two known subgroups. Here, we refer to Group 3 clusters as Group 3a (G3a) and Group 3b (G3b), and SHH clusters as SHHa and SHHb (Figure 2A, Table S1). Including WNT samples in any of the clustering approaches on the proteomic data revealed the same five subsets with an additional sixth group for WNT samples (Figure S3). Performing pairwise statistical comparisons, we found 4,365 proteins and 2,642 phosphopeptides that differed significantly between these subsets of patients (FDR < 0.01, ANOVA; Table S2).

Proteome suggests post-transcriptional heterogeneity within SHH medulloblastoma

The SHHa and SHHb subsets that emerged (Figures 2A, S1-S3) are distinct from the age-dependent subtypes found in DNA methylation data (Cavalli et al., 2017; Kool et al., 2014; Northcott et al., 2011; Schwalbe et al., 2017). All but one of the adult samples were found in SHHa, while the pediatric SHH samples were split evenly between SHHa and SHHb (Figures 2B, 3, S3, Table S1). We identified 510 proteins that differed between the SHHa and SHHb using ANOVA (FDR < 0.005). Proteins higher in SHHa were associated with mRNA processing, splicing, and transcription, as well as the MYC pathway, chromatin remodeling, and DNA repair (Figure 3A, Table S3). In contrast, proteins with higher levels in SHHb were linked to neuronal and neurotransmitter-like activity, including CD47, an anti-phagocytic cell surface ligand (Figure S4). Many proteins in the glutamatergic synaptic pathway were elevated in SHHb, including glutamate, calcium, and MAPK/ERK signaling (Figures 3A, 3B, Table S4). SHHb samples consistently clustered closer to Group 4 samples than to SHHa samples (Figures 2A, 3A, S1-S3). Despite the differences in neuronal-like gene sets between SHHa and SHHb, there were no differences in histology (Figure 3C).

As no transcripts (RNA-seq) differed significantly (FDR < 0.05, ANOVA) between SHHa and SHHb, we asked whether the subsets differed in post-transcriptional regulation. We compared the Spearman correlations for approximately 8,700 mRNA-protein pairs in every patient sample (Figure 3D). All clusters of samples except SHHb have a median Spearman correlation near 0.5, consistent with studies in other systems (Mertins et al., 2016; Zhang et al., 2014). However, the median correlation for SHHb was 0.38 (p = 0.012, compared to SHHa; Mann-Whitney U test), suggesting that SHHa and SHHb differ in translation and/or proteostasis.

We next asked if genetic lesions might account for the observed differences between SHHa and SHHb. Gains of chromosome 3q are characteristic of SHH medulloblastoma (Cho et al., 2011; Kool et al., 2012; Taylor et al., 2012), and in our cohort they occur in several SHHa samples, and do not occur in SHHb samples (Figure 3C). The neural progenitor regulator SRY-box 2 gene (*SOX2)* lies within this region, and its protein levels and phosphorylation events (T7, S18, and T26) were all increased in SHHa samples compared to SHHb (Figures 3C, S4A) (Archer et al., 2011). In contrast, CD47 protein levels were significantly elevated in SHHb, while mRNA levels did not differ between the groups (Figure S4B). Mutations in *PTCH1*, *PRKAR1A*, and in the *TERT* promoter (Figure 3C) occurred only in SHHa. In contrast, other SHH pathway activating alterations were found in both SHHa and SHHb, which suggests that there are relatively few differences in the genetic lesions of these SHH subsets.

Post-translational modifications of MYC in Group 3 tumors are predictive of patient outcome

Clustering of the proteomic data identified two subsets of Group 3 samples, which we refer to as G3a and G3b (Figures 2, S1-S3). First, we sought to understand how these clusters related to the subgroups that had been previously identified using much larger cohorts. All G3a samples were assigned to subtype II that was defined by Northcott *et al.* (2017), while the G3b samples were assigned to the Northcott *et al.* (2017) subtypes I, III, IV, and VII (Figure 2B, Table S1). The transcription-based classifier of Cho *et al.* (2011) assigned the G3a samples to the MYC-activated c1 subtype. Two G3b samples and one Group 4 sample also were classified as c1 (Figure 2B, Table S1). The remaining G3b samples were assigned to c4 and c5. These results suggest that the proteomic features associated with G3a likely represent the MYC-activated form of medulloblastoma and that the proteomic data for G3b samples represent the known Group 3/4 continuum. Indeed, several proteomic signatures in G3a associated with MYC activation including significant differences in ribosomal proteins and proteins related to ribosome assembly, mitochondrial ribosomal proteins, and proteins involved in transcription (Figure S5A) (Morrish and Hockenbery, 2014; Staal et al., 2015).

To identify the sources of MYC activation in G3a, we investigated the MYC events in all our data (Figure 4A). While *MYC* amplification is a "hallmark" of MYC-activated medulloblastoma, it does not occur in every tumor of this type (Cho et al., 2011; Northcott et al., 2017). Here, only two G3a tumors have a *MYC* amplification. However, all G3a samples have increased post-translational modifications of MYC at multiple sites (Figures 4A, S5B). Acetylation of lysine 148 (K148) was pronounced, as was increased phosphorylation of serine 71 (S71) and a serine at either position 62 or 64 (the ambiguity of these nearby sites cannot be resolved in the spectra). Our data also revealed peptides that are simultaneously phosphorylated at both serine (S62) and threonine 58 (T58). Peptides phosphorylated on T58 and S62 are particularly informative for MYC activity, as these sites regulate MYC half-life and transcriptional activity (Arnold et al., 2009; Wang et al., 2011). Considering known MYC regulators, our data showed significantly increased protein levels of the B55a subunit of PP2A (encoded by *PPP2R2A*, FDR = 0.0085, ANOVA), and the deubiquitinating enzymes USP28 (significant with FDR = 0.018, ANOVA) and USP36 (not significant with FDR = 0.084, ANOVA) in G3a compared to G3b (Figure S5C). By contrast, we did not

observe differing levels of PP2A-B56 or the phosphatase inhibitors SET or CIP2A (Figure S5C).

To examine the localization of active forms of MYC, we stained formalin-fixed and paraffin-embedded (FFPE) slides of the tumors for MYC phosphorylated at S62 or T58 (Figure 4B). Cell mean immunofluorescence density for phosphorylation of residues S62 and T58 MYC (pS62 and pT58) was significantly higher in G3a compared to G3b (Figure 4C), while Group 4 tumors showed lower levels of signal. Localization of pS62 MYC was primarily in the nucleus with exclusion from the nucleolus (Figure 4B). In contrast, pT58 MYC localized mainly in the cytoplasm for G3a and G3b tumors. Elevated pS62 and pT58 MYC suggest a breakdown in the canonical pathway of MYC degradation (Farrell and Sears, 2014), consistent with the increased expression of some kinases upstream of S62 phosphorylation as well as increased expression of deubiquitinating enzymes USP36 and USP28.

We sought to understand whether the observed MYC modifications had clinical implications. Indeed, G3a and G3b patients differed dramatically in rates of five-year progression free (PFS) and overall survival (OS), but these differences were not statistically significant, possibly due to the small sample size (Figure 4D). To extend our analysis to a larger cohort, we built a binary classifier based on single sample gene set enrichment analysis (Barbie et al., 2009) to distinguish G3a and G3b using only transcriptional data. We then applied it to the c1 and c5 samples from Cho *et al.* (2011) to give them G3a/G3b labels (Figure S5D, Table S1). Surprisingly, while we had expected c1 samples from the original Cho cohort to be assigned to G3a and c5 to G3b, the classifier assigned approximately one quarter of the samples to the other subtype. The combined cohort using proteomic labels performed better at predicting PFS and OS (Figure 4D, S5E). The difference in PFS using c1 and c5 labels (Figure 2B and Table S1) on the combined cohort did not reach statistical significance ($p = 0.098$), while the difference using G3a vs. G3b classification did ($p = 0.002$). The newly defined sets of patients had the same median age (5 years), ruling out that variable as a potential trivial explanation for the improved predictor. Taken together, these data demonstrate that the proteomic distinctions provide a strong signal of clinical relevance.

Medulloblastoma subgroups differ in activity of kinases

We next investigated kinase signaling in medulloblastoma. Potential kinases of the phosphopeptides in each subset of tumors were identified by leveraging kinase specificity data from PhosphoSitePlus (Hornbeck et al., 2015) and Scansite (Obenauer et al., 2003). Filtering this list based on the protein and phosphorylation levels of the kinases in our data, we identified the 38 kinases listed in Figure 5 (Table S5). PRKDC, GSK3B, and CDK5 were significant in both our analyses. PRKDC is important for repair of DNA double-stranded breaks through non-homologous end-joining (Ma et al., 2004) and downstream targets of PRKDC were consistently elevated in Group 3 and WNT samples (Figure 5, Scansite). GSK3B is a promising therapeutic target. It is overexpressed in many cancers (McCubrey et al., 2014) and its substrates have increased phosphorylation in the Group 4 and SHHb samples. CDK5 has been associated with oncogenesis and resistance to cancer therapies (Pozo and Bibb, 2016) and its targets had increased phosphorylation in Group 4. CDK5 phosphorylates MYC at S62, much like CDK1 and CDK7, which were fittingly associated

with G3a. AURKB was also linked to this disease group. Diaz *et al.* (2015) previously treated MYC overexpressing medulloblastoma cell lines and orthotopic xenografts with an AURKB inhibitor, which led to growth impairment and induction of apoptosis in the cell lines, and inhibited intracranial growth and prolonged survival in mice. Many kinases were predicted to be active in SHHa, and this was the only subgroup with consistent phosphorylation of sites targeted by ATM (PhosphoSitePlus) and PIK3R1 (Scansite). G3b was unusual, as only one kinase, EEF2K, was associated with this group in our analyses. Overall, these analyses highlight potential upstream kinases that may be driving some of the molecular differences between medulloblastoma subgroups and that represent possible novel targets to explore for future therapeutic gains.

MYC-active medulloblastoma cell lines have phosphorylated MYC and PRKDC

To test the finding that pS62 MYC is enriched in Group 3 MYC-active medulloblastoma, we measured MYC expression by Western blot in medulloblastoma cell lines reported to be *MYC*-amplified (D425, D458 and D556) compared to two that are not (DAOY and D283) (Bigner et al., 1990). *MYC*-amplified lines were highly enriched for MYC protein and pS62 MYC, while DAOY and D283 were not (Figure 6A-C). Consistent with the kinome analysis, pS2056 PRKDC correlates with pS62 MYC and pT58/pS62 MYC in these lines (Figures 6A-C). To explore possible functional consequences of this finding, we first examined the localization of PRKDC and MYC. Immunofluorescence identified that pS62 MYC and pS2056 PRKDC co-localize in the nucleus (Figure 6D).

We next examined the role of PRKDC in *MYC*-amplified medulloblastoma cell lines. PRKDC inhibitors are radio-sensitizing agents, but may have limited cytotoxic activity by themselves (Ciszewski et al., 2014; Sunada et al., 2016). We find that that the PRKDC inhibitor NU7441 preferentially sensitizes MYC-active cell line D458 to radiation (Figures 6E-F). Irradiation reduced the IC_{50} of the *MYC*-amplified D458 from 28 μM to 2.7 μM (Figure 6G), but did not change the IC_{50} of NU7441 in DAOY. These data suggest that cell lines with pS62 MYC depend on PRKDC activity for survival in response to DNA damage, and that PRKDC inhibition may radio-sensitize MYC-active tumors.

Integrative modeling

To search for common patterns in the genomic, proteomic, and phospho-proteomic data, we adopted an integrated modeling approach (Figure 7). Omics Integrator (Tuncbag et al., 2016) searches a vast network of physical interactions for sets of proteins and genes from disparate 'omic data that are likely to represent pathways altered in a disease process. Applied to signals that differ significantly between G3a and G3b, it identified coherent proteomic changes in proteins that physically associate with MYC, and up-regulation of known MYC transcriptional targets (Figure 7A). In addition, the networks highlighted ribosomal, mitochondrial and cell-cycle regulatory proteins. A similar analysis for SHH (Figure 7B) identified calcium, glutamate and Ras signaling pathways that were upregulated in SHHb. These networks integrated disparate data types including: mutated genes (*TP53*, *PIK3R1*, and *NOTCH1*) and kinases with genomic/proteomic/PTM changes (CAMKs (CAMK2A/G, CAMK4), protein kinase A and C (PRKACB, PRKCA), ERK2 (MAPK1), ribosomal protein kinase 6 (RPSK6KA1/2), glycogen synthase kinase B (GSK3B), PI3Ks (PIK3R1, PIK3CA),

DGKI, and CDK2). The networks also supported the role of several predicted kinases (PIK3R1, CAMK2G, GSK3B, and MAPK1).

Discussion

Molecular subgroups identified using mRNA expression and DNA methylation array data have recently been accepted by the WHO as the international standard for medulloblastoma diagnosis (Louis et al., 2016a, 2016b). Here we show that, at the highest level, proteomic and phospho-proteomic data sets revealed similar subgroup assignments to these consensus subgroups. This finding contrasts with global proteomic studies in breast and colon cancer where molecular subgroups were not as consistently durable across data types, and subgroup compositions are dependent on the data type used (Mertins et al., 2016; Zhang et al., 2014). The consistency of the medulloblastoma subgroups may in part be due to the distinct developmental states of the cell of origin of medulloblastoma subgroups, as well as the subgroup-specific driver genetic events (Gibson et al., 2010; Northcott et al., 2012).

Our data identify heterogeneous molecular mechanisms within the subgroups that are not evident in the transcriptome or genome. The two clusters of SHH samples revealed by our proteomic data (SHHa and SHHb) reflect a different distinction from the known age-based split in SHH samples (Cavalli et al., 2017; Kool et al., 2014; Schwalbe et al., 2017). Adult patients in our cohort (ages 23–35) predominantly clustered with SHHa, but the infant and childhood patients were spread across both subsets. The different signals that emerged from DNA methylation and proteomics may reflect the differing sensitivity of these methods to specific biological processes. DNA methylation data likely reflect the developmental state of the cells of origin at the onset of oncogenesis (Horvath et al., 2015; Lu et al., 2016). By contrast, the proteomic data will be strongly affected by post-transcriptional changes including RNA stability, protein stability, translational regulation, and signaling pathways.

The proteomic data we gathered from SHH tumors have important clinical implications. SHHa samples had expression signatures and molecular alterations including *PTCH1* mutations that are consistent with activation of the canonical SHH pathway. SHHb tumors also had SHH pathway activating mutations, but also were enriched for pathways typically associated with Group 4, such as glutamate, calcium, and Ras signaling. For example, we saw very consistent increases in many of the proteins associated with the glutamatergic synapse. While these pathways are known features of some gliomas and Group 4 medulloblastoma (Arcella et al., 2005; Cavalli et al., 2017), they are not typically associated with SHH medulloblastoma. SHHb-like patients may therefore also benefit from any future therapies developed for Group 4 tumors. Proteomic data identifies one potential candidate therapeutic target: CD47 protein is enriched in Group 4 and SHHb tumors. CD47 is a membrane protein involved in several processes, including vesicle-mediated transport, and is an anti-phagocytic cell surface ligand (Brown and Frazier, 2001; Jaiswal et al., 2009). Currently, an anti-CD47 antibody is being tested for efficacy in Group 3 medulloblastoma (Gholamin et al., 2017). While *CD47* RNA levels are not significantly different between the SHH groups (p = 0.06, t-test), the mean CD47 protein level for SHHb samples is much higher than the mean for SHHa and Group 3 samples. These data suggest that anti-CD47 therapies may be particularly effective on SHHb tumors. The discrepancy between protein

and mRNA levels in CD47 is not unique, as the correlation of global protein levels with mRNA expression was significantly lower in SHHb samples. It will be important for future studies to examine whether there exist fundamental differences in the efficiency of protein translation or in the stability of proteins in SHHb tumors. These observations highlight the importance of proteomic studies for discovery of biomarkers.

Group 3 medulloblastoma are clinically diverse and it has been clear that MYC activation has important therapeutic consequences. Previously, we demonstrated that transcriptional profiles can distinguish low- and high-risk Group 3 patients (Cho et al., 2011; Tamayo et al., 2011). More recent studies have proposed additional subtypes of Group 3. Cavalli *et al.* (2017) proposed three subtypes, and Northcott *et al.* (2017) identified eight subtypes of Group 3 and Group 4. However, genomic, epigenomic, and transcriptional data do not directly measure activation of MYC, which can occur through several mechanisms. In breast cancer, for example, more than 40% of tumors show increased levels of MYC protein, but the fraction with increased *MYC* mRNA ranges from 22–35% and only 13–22% have an amplification of the *MYC* locus (Chen and Olopade, 2008). In breast cancer, higher MYC protein levels without genomic amplification have been explained by elevated levels of pS62, a phosphosite associated with more stable and transcriptionally active MYC (Janghorban et al., 2014). Using the direct measurement of MYC post-translational modifications, we were able to refine the transcriptional signature of MYC-activated Group 3 tumors. It is clear from the pathway-level analysis that G3a tumors have higher levels of MYC activity compared to those in the G3b cluster. Consistent with these data, we see pS62 on MYC in the proteomic data and by staining both of FFPE tumor sections and medulloblastoma cell lines; and we found that pS62 MYC was primarily expressed in the nucleus. More work will be needed to understand the functional consequences of the observed post-translational modifications.

Through systematic analysis of the phospho-proteomic data, we have identified several kinases that should be studied further to understand their therapeutic implications. PRKDC was predicted by our kinome analysis and its levels were elevated in many Group 3 and WNT samples. While there is no prior evidence for PRKDC's role in medulloblastoma, *in vitro* experiments with various tumor cell lines show that PRKDC promotes MYC stability (An et al., 2008). PRKDC and MYC are known to function together in the presence of DNA damage (Cui et al., 2015). Notably, we have shown that in the presence of high levels of endogenous MYC activity in medulloblastoma cell lines, PRKDC inhibition functions as a radiation sensitizer, but not in cells with low MYC levels. Indeed, our data suggest that the radio sensitizing effects of PRKDC inhibitors may be dependent on MYC status, and furthermore that PRKDC inhibition may serve as radiation sensitizer of MYC-active G3a medulloblastoma in clinical trials.

In conclusion, our results show that quantitative mass spectrometry-based proteomics reveals molecular mechanisms within medulloblastoma subgroups that are not evident through analysis of genome, epigenome, or transcriptome. Protein expression and post-translational modifications represent the functional state of the cancer cells, a reflection of the influence that somatic mutations and other genetic and epigenetic alterations have to alter the cellular state during progression from normal to a cancerous state. Kinome analysis is a particularly

sensitive method to identify specific kinases for therapeutic targeting. Finally, differential modification of proteins through post-translational modifications offers new biomarkers for specific medulloblastoma subtypes. Our integrative exploration of medulloblastoma furthermore provides the clinical and research communities with a wealth of data that may help advance strategies for patient selection and treatments for this devastating disease.

It has come to our attention that a parallel study by Forget *et al.* (2018) also identifies two subgroups of SHH medulloblastoma. Their pathway analysis also finds SHHa samples to be enriched for DNA replication processes, and SHHb to be enriched for neuronal and neurotransmitter genes.

STAR Methods

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, E.F. (fraenkel-admin@mit.edu).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Patient samples—Primary medulloblastoma patient samples, including FFPE slides, were obtained with informed consent according to the International Cancer Genome Consortium (ICGC) guidelines as approved by the Ethics Committee of the Medical Faculty at Heidelberg University, and as approved by the institutional review board of contributing center Nikolay Nilovich Burdenko Neurosurgical Institute in Moscow. All samples were de-identified. Tumor samples of 50 mg were freeze-fractured using Covaris cryoPREP CP02 at setting "impact level 4". The pulverized samples were aliquoted for the downstream methods.

Cell Lines—Medulloblastoma cell lines D425, D458 and D556 were a kind gift from Dr. Darell Bigner (Duke University). DAOY and D283 were obtained from American Tissue Culture Collection.

METHOD DETAILS

Proteomic profiling—The global proteome and phospho-proteome were processed according to adapted protocols from our previous studies (Huang et al., 2017; Mertins et al., 2016). In brief, cryo-pulverized tumor tissue from each patient was lysed at 4°C using 8M urea lysis buffer. Extracted proteins were reduced using dithiothreitol and alkylated with iodoacetamide before digestion using LysC for two hours followed with trypsin overnight. Both digestion steps were performed at a 1:50 enzyme:protein ratio. For relative quantification of the global proteome and phospho-proteome by liquid chromatography tandem mass spectrometry (LC-MS/MS), 400 μg per patient, as measured on protein level (BCA protein concentration determination kit; before digestion) was labeled with 10-plexing tandem mass tags (TMT-10; Thermo Scientific) following the manufactures instructions. All 45 patients were run in 5 total TMT-10 plexes, with each plex including 9 patient samples and an internal reference sample. Samples were assigned to plexes in a semi-randomized manner, such that consensus subgroups were split across TMT-10 plexes (assignments in

Table S1). The internal reference sample was composed of equal amounts of peptide material from 40 of the 45 patients, representing all subgroups, and was included in each TMT10-plex to provide a common standard for precise relative quantitation. Isobarically-labeled peptides were combined and fractionated using high-pH reversed phase chromatography into 24 fractions. From each fraction, 5% of the material was evaluated for its proteomic content. The remaining 95% material was combined into 12 fractions which were each enriched for phosphopeptides using immobilized metal affinity chromatography (IMAC) (Mertins et al., 2013). The flow-throughs after IMAC enrichment were collected and further concatenated into 4 fractions that were each enriched using antibodies for acetylated peptides (see Key Resource Table). In parallel to the global proteome, phospho-proteome, and acetylome, an additional 500ug of TMT-labeled peptides per patient were enriched for phosphotyrosine peptides using phospho-tyrosine antibodies (see Key Resource Table) and analyzed as a single fraction on the mass spectrometer. All proteomic based data were collected on a Lumos mass spectrometer (Thermo Fisher Scientific) and the resulting spectra were searched using Spectrum Mill (Agilent, version 12.212). All mass spectra contributing to this study can be downloaded in the original instrument vendor format from the MassIVE online repository (MSV000082644).

Sequencing and DNA Methylation Array Data Collection—Whole genome sequencing, RNA expression, and DNA methylation data sets reported here are previously published in Northcott *et al.* (2017) (European Genome-phenome Archive EGAS00001001953). Consensus subgroup assignments (Figure 1) were provided by DKFZ, and methylation-based subgroups were assigned previously in Northcott *et al.* (2017).

Western blots—Cells were cultured as previously described (Weeraratne et al., 2012). For Western blots, 1 million cells were plated in 10 cm dishes and harvested after 48 hours. Proteins were normalized using Pierce BCA Protein Assay Kit and 50 μg protein was loaded per well into NuPAGE Novex gels with Bolt running buffer (see Key Resource Table for specifics). Proteins were transferred using iBlot2 system. Blots were probed for primary antibodies and visualized using the Licor Odyssey system on the Odyssey CLx Infrared Imaging System according to manufacturer's directions. Antibodies dilutions for Western blots are listed on Key Resource Table. Antibody staining was quantified using Image Studio Lite for Western blots. Densitometry was performed by comparing raw densitometry for each antibody to actin on unmodified images.

Immunofluorescence of cell lines—Immunocytochemistry was performed as previously reported (Weeraratne et al., 2012) with the following modifications: cells were plated at 30,000 cells/well in 500 μl of media on glass coverslips. Slides were imaged using Zeiss 710 Confocal Microscope in the IDDRC and analyzed using Fiji Image J. All antibodies were used at 1:50 concentration for immunofluorescence.

Drug dose response assay—Adherent DAOY cell lines were plated at 2,500 cells per well and suspended D458 were plated at 20,000 cells per well in 96-well plates in 75 μL of culture media. Each sample was plated in sextuplicate. NU7441 was added 24 hours after plating at 4x concentrations in 25 μL. CellTiter-Glo Luminescent Cell Viability Assay was

used to measure viability 18 hours after drug addition according to manufacturer's directions in white opaque plates. Luminescence data was measured by EnSight Multimode Plate Reader using Kaleido 1.2 software with 0.1 second measurement time. For irradiation, the cells were exposed to 200 rads of gamma radiation and assayed for cell viability 5 hours later.

Immunofluorescence staining of FFPE slides—Antigen retrieval was achieved by pressure-cooking in citrate buffer pH 6 (Sigma) for 10 minutes. Antibodies used for staining are as follows (details in Key Resource Table): rabbit polyclonal c-Myc S62 specific phospho-antibody 1:25 (Zhang et al., 2012) and rabbit polyclonal c-Myc pT58 antibody 1:50 (Applied Biological Material) incubated overnight at 4°C. Secondary antibody was Alexa Fluor 594 1:500 (Invitrogen) and DAPI at 1:5000 (Sigma) incubated for 1 hour at room temperature. ProLong Gold mounting media (LifeTech) was used and allowed to cure for at least 24 hours. Images were taken with a Hamamatsu digital camera (Japan) mounted on a Leica fluorescence microscope (Wetzlar, Germany) at 40x. Representative images (Figure 4B) were acquired at 63x on a Zeiss LSM 880 laser-scanning confocal microscope (Germany).

QUANTIFICATION AND STATISTICAL ANALYSIS

Processing of DNA Methylation Array Data—We used the *minfi* R library (Aryee et al., 2014) to process the IDAT files into quantile normalized beta values. The probes' beta values were then collapsed to gene symbols using the means of geneassociated probes.

Processing of genomics data—This section describes the (re)processing of genomics data (RNA-seq, WGS) for the medulloblastoma cohort that is the subject of this study. All genomics data were available prior to this study and published elsewhere (Northcott et al., 2017). Details about sequencing protocols can be found in the corresponding publications. We decided to re-process all data using the latest best-practice pipelines developed at the Cancer Genome Analysis (CGA) group of the Broad Institute.

Processing of RNA-seq data—Bam files were unaligned and converted to FASTQ using Picard (http://broadinstitute.github.io/picard/). All further RNA-seq data processing described below was conducted in FireCloud, a cloud-based computing environment developed and maintained at the Broad Institute. Briefly, RNA-seq reads (50 bp) were aligned to GRCh37 (UCSC hg19) genome assembly using STAR aligner (Dobin et al., 2013). For each sample we assessed QC metrics using RNA-SeQC (DeLuca et al., 2012). Transcript expression was quantified as Transcripts Per Kilobase Million (TPM) using RSEM (Li and Dewey, 2011).

Processing of Affymetrix expression array data—Expression array data was quantile normalized using the preprocessCore R library (function "normalize.quantiles").

Processing of WGS data—The processing of 44 WGS tumor-normal pairs described below was accomplished by chaining together modules implemented in GDAC Firehose (http://gdac.broadinstitute.org/). Somatic variant calling modules were based on the Genome

Analysis Tool Kit v3 (GATK) following 'GATK Best Practices' workflows for variant discovery (DePristo et al., 2011).

Paired-end (PE) sequencing reads (100 bp) in FASTQ format were aligned to GRCh37 (UCSC hg19) genome assembly using BWAMem. Base quality score recalibration (BQSR) to correct systematic errors made by the sequencer was done using GATK's 'BaseRecalibrator' and 'PrintReads' programs. Local realignment of insertions and deletions (indels) to correct alignment artifacts was performed with GATK's 'IndelRealigner' program. Cross-sample contamination was assessed by GATK 'ContEst' program.

Mutation calling. Somatic single nucleotide variants (SNVs) and indels were called using Mutec2. Resulting VCF files were annotated and converted to MAF format by Oncotator (Ramos et al., 2015). Oxidative artifacts contributing to SNV calls were assessed by D-ToxoG. SNVs and indels were further filtered for commonly observed germline variants using a panel of normal (PoN) (Costello et al., 2013).

Copy number calling. To derive somatic copy number variants, we used a GATK v4 workflow that comprised four steps: 1) Proportional coverage per read group was calculated by the 'CalculateTargetCoverage' program. 2) A panel of normals (PoN) was created from the normal samples which is meant to encapsulate sequencing noise and common germline variants. The program 'CombineReadCounts' combined proportional read counts from all normal samples into a single file. The PoN file stores information like the median coverage target and was generated using 'CreatePanelOfNormals' using Principle Component Analysis (PCA) to calculate systematic noise. 3) Normalization of tumor coverage by PoN target medians and PoN principle components using 'NormalizeSomaticReadCounts'. Resulting data were log2 transformed. 4) Segmentation of groups of contiguous targets with the same copy ratio using the 'PerformSegmentation' program.

To derive gene-centric copy number variant calls we used GISTIC2.0. Contiguous gene copy number ratios were corrected for diploidy; a value of zero indicates the presence of two gene copies.

Normalization of proteomics data—Relative expression data derived from proteomics profiling (proteome, phospho-proteome, acetylome) were separately normalized by sample using robust z-scores (Z_r): $Z_r = (x - M)/MAD$. In this expression, M is the median expression of the sample; and MAD is the median absolute deviation of the sample. To allow for a better comparison across data types, WNT samples were excluded from these analyses as they lacked all but the proteomic data, and there was insufficient high-quality tissue to perform the additional assays.

Normalization of RNA-seq expression data—TPM values were first converted to relative scale by median-row normalization. Resulting ratios were then transformed to robust z-scores as described above (Table S2).

Quantification of Immunofluorescence staining—Mean immunofluorescence densities were generated using OpenLab 5.5 software (Improvision) by using the ROI tool to

quantify 50 to 100 nuclei in 3 to 4 fields of view at 40x magnification. Scatter and bar plots were generated using GraphPad Prism. Error bars represent SEM and significance reported is from two-tailed unpaired t-tests.

Quantification of Western Blots—Raw densitometry for each antibody was normalized to actin using Image Studio Lite. All quantification was performed in at least triplicate on original unmodified blots. Normalized signals were compared between cell lines by arbitrarily normalizing samples against the DAOY line. One-way ANOVA with a Dunnett multiple comparison test was performed to compare the normalized signal for each antibody across the five cells lines using GraphPad Prism 5. Error bars represent mean normalized signal ± SEM.

Dose response curve—Luminescence for each well was measured and the mean signal for each condition ± SEM plotted. Because adherent (DAOY) and suspended (D458) cells were compared, we normalized mean luminescence to 100% for the non-irradiated treatment group for each cell line. A dose-response curve and the IC_{50} were determined by non-linear regression with GraphPad Prism.

Survival Curves—Group 3 samples were divided into G3a and G3b groups. G3a consists of the samples in our cohort assigned to G3a and the samples from the Cho *et al.* (2011) cohort predicted to be G3a. The same scheme applies to G3b. Kaplan-Meier curves were generated for each of these groups, for both overall survival (OS) and progression-free survival (PFS). All Kaplan-Meier curves and log rank test p values were generated with the lifelines Python package (Davidson-Pilon et al., 2017). Samples whose death event were not observed within the 10 years following initial diagnosis were right-censored for Figure 4D, and the full outcome follow up Kaplan-Meier plots are available in Figure S5E.

Group 3 Cohort Expansion—To expand our PFS and OS analysis for G3a and G3b, we developed a Group 3 Cohort Expansion predictor (https://github.com/ckmah/ archer2018_g3_cohort_expansion/) following the methods in Cho *et al.* (2011). From the Cho *et al.* (2011) cohort, we used the c1 and c5 samples which are most like Group 3. Since only array data were available for the Cho samples, we also used array data for our current Group 3 cohort to better normalize the two data sets. We projected the expression data for both cohorts into "gene set space" using a single sample version of GSEA (ssGSEA), using the Hallmarks (H), Curated Gene Sets (C2), Motif Gene Sets (C3), and Oncogenic Gene Sets (C6) collections from the MSigDB. Using the G3a, G3b labels for the current cohort, we determined the 10 most differentially enriched gene sets for each subtype (Figure S5D, main text) according to the Information Coefficient (IC) as defined in Kim *et al.* (2016). These top sets were used as features to train a Bayesian cumulative logodds predictor as previously described (Tamayo et al., 2011). The predictor was applied to the projected Cho *et al.* (2011) c1 and c5 samples to assign G3a and G3b labels and those labels were used to construct a combined cohort PFS and OS analysis. We used the Cho *et al.* (2011) labels for the current cohort as given in Figure 2B of the main text to construct a combined c1/c5 labeled cohort for PFS and OS analysis. This analysis did not include sample MB166 and MB278 from G3b which were labeled c4. The c1/c5 labeling of the current cohort used the

classifier developed for Cho *et al.* (2011), with two minor modifications: (1) we restricted the analysis to genes contained in both data sets, and (2) the IPA_KCNIP2_DN gene set used as a feature for c5 in Cho *et al.* (2011) was replaced by BIOCARTA_PS1_PATHWAY. The classifier's group assignments are provided in Table S1.

Glutamate Pathway modeling—For Figure 3B, the gene sets were collected from the Kyoto Encyclopedia of Genes and Genomes (KEGG). Protein levels were summarized using the median log2-normalized protein level for a gene for all samples in a group (SHHa or SHHb). When multiple genes are in a gene set, the average of those gene values is shown. We only included genes with data for at least 75% of samples from both groups. A summary of all the genes included in Figure 3B is included in Table S4.

Consensus Clustering—We used the ConsensusClusterPlus R library (Wilkerson and Hayes, 2010) to perform consensus clustering on our data sets. For each data set we varied k, the number of clusters, from 2 to 9 and ran the algorithm with 1,000 subsamples for all combinations of two clustering methods (hierarchical clustering and k-means) and three distance metrics (Euclidean, Spearman, Pearson). Consensus heatmaps are available in the Supplement (Figure S1). Consensus clustering of proteomic data indicated that this data type produced highly consistent clusters with k = 5. Runs with k < 5 showed a clear substructure within the clusters identified. Notably, for all clusterings with k < 5, the SHHb samples clustered with Group 4 samples rather than the SHHa group. Runs with k > 5 largely included very small clusters or samples that showed pronounced cluster promiscuity. We also calculated PAC (proportion of ambiguously clustered pairs) scores (Senbabaoglu et al., 2014) for our consensus clustering runs (Figure S1) and found confirmation that, for proteomics data, k = 5 was the optimal number of clusters when trying to avoid clusters with three or fewer samples. Using the PAC scores we also determined that hierarchical clustering and Pearson-based distance was the best parameter setting for this *k*. With these settings, one sample, MB247, frequently shifted between G3a and G3b. We decided to assign it to G3a based on results from other unsupervised methods such as PCA.

Dimensionality Reduction—Principal component analysis (PCA) and t-distributed stochastic neighbor embedding (tSNE) were employed to display our high-dimensional data sets in two dimensions. We used the R function "princomp" in the "stats" library for PCA, and the "Rtsne" package for tSNE. Running Rtsne, we used the function's default parameters along with the maximum perplexity the program would allow for (perplexity = (number of samples – 1) / 3).

Differential analysis—We used two types of analyses to find differential features in our proteomic data sets. One strategy sought to extract features specific to a single disease group in comparison to the rest of the samples; the other approach selected features that differ between pairs of groups. We performed univariate t-tests (using "t.test" in R) to find features between each disease group and all the other samples. For pairwise comparisons across all groups, we performed ANOVA along with R's "TukeyHSD" method to calculate p values. P values from both methods were corrected for multiple hypothesis testing using the Benjamini-Hochberg/FDR method implemented in R's "p.adjust" function. Due to concerns

associated with the small number of samples (in particular for comparisons involving SHH samples), we aimed to focus the comparison on the most homogeneous groups possible. We therefore withheld sample MB136 from these statistical tests, as it clustered stably with SHHb in the proteomic data but showed similarities with SHHa in other data types.

We are aware that the small sizes of our cohort's disease groups present challenges in the selection of statistical tests and estimation of significance of results. We selected the methods mentioned above as they were able to identify differential features with reasonable confidence and allowed us to prioritize features to feed into downstream analysis methods (e.g. Omics Integrator). These downstream methods served as a second filter and none of our conclusions are directly based on the results of these tests. For statistics on individual univariate comparisons we used a Mann Whitney U test (these instances are indicated as such in the main text). Across the different testing methods, we considered FDR-adjusted p values of less than 0.05 as significant, but chose more stringent thresholds for certain analyses to limit the number of features. Table S2 contains a complete list of proteins and peptides that are significantly differential between our proteomic groups (for FDR < 0.01, ANOVA). In this table, we also indicate differential kinases (FDR < 0.005).

Global Correlation Analysis—To correlate proteomic and RNA-seq data, we considered only those data with no missing values and identified matching proteins and transcripts based on gene symbols. In doing so, splice isoforms were collapsed by their means. For each sample, we then calculated the Spearman correlation for the 8,674 proteins and their corresponding mRNAs.

Functional Annotation of Data Sets—To project data sets into the space of gene sets representing functional information, we performed hypergeometric overlap tests (background size context dependent) and used an in-house implementation of single sample Gene Set Enrichment Analysis (ssGSEA) (Barbie et al., 2009; Subramanian et al., 2005). For the latter, we used the following parameter settings: rank-based sample scoring ("sample.norm.type": ranks), area under curve statistic to calculate scores ("statistic": area.under.RES), default weight (0.75), 500 permutations, z-score pre-processing ("correl.type": z.score), and normalized enrichment score output ("NES").

We used the gene sets available in the Molecular Signature Database (MSigDB v6.0) (Liberzon, 2014; Liberzon et al., 2015), primarily focusing on canonical pathways (C2CP), GO terms (C5), and Hallmark sets (H). P values from these methods were corrected for multiple hypothesis correction using the Benjamini-Hochberg/FDR correction implemented in R's "p.adjust" function. FDR-adjusted p values less than 0.05 were considered significant.

Integrative Network Analysis—We used the latest version of Omics Integrator 2 (OI2 v0.2.24, https://github.com/fraenkellab/OmicsIntegrator2) to construct disease-focused, integrative networks using proteomic and genomic data (Tuncbag et al., 2016). Omics Integrator begins by mapping a set of proteins of interest onto the nodes in a network of physical interactions ("interactome") among proteins. The nodes are assigned "prizes", reflecting their importance (see below for details). The interactions were derived from public databases using the iRefIndex v14 collection of interactions (Razick et al., 2008). Each

interaction is associated with a cost that is lowest for the most reliable interactions (calculated as 1 minus the edge score provided by iRefIndex). We also added previously published site-protein (cost: 0.25) and sitekinase (as published in PhosphoSitePlus, (Hornbeck et al., 2015)) edges (cost: 0.4) to the interactome to allow the algorithm to find site-kinase interactions in the solution that are not included in the iRefIndex interactome. The algorithm seeks to identify subnetworks that contain many disease relevant nodes (based on prizes) while still avoiding using too many low confidence edges (based on costs). For more details about this method, see Tuncbag *et al.* (2016).

We tailored the selection and definition of prizes to the type of network we sought to detect. We created networks focusing on the differences between G3a and G3b ("G3 network"), as well as SHHa and SHHb ("SHH network"). Input nodes were selected based on differential protein or peptide levels for the relevant sample group comparison in each network. For the G3 network, proteomic features needed to pass the same FDR threshold (FDR < 0.05) in each data set. For the SHH network, proteins and phosphopeptides in pSTY needed to pass a more stringent threshold (FDR < 0.005) due to the large number of differential features, while the threshold for acetylated peptides and phospho-tyrosine peptides was set to 0.05. Prizes for nodes based on proteomic data sets were calculated as the absolute value of the log2-based fold changes (based on the sample groups' means) from pairwise comparisons. Genomic events were each assigned a fixed prize at an arbitrary value of 2.5 and each CNA and mutation found in at least one sample in the relevant group was given a prize. This value was used as we wanted to lend strong weight to genomic alterations, but not have them outweigh strong proteomic signals. We therefore opted for a value between the median and the 3rd quartile in the proteomic prize distributions.

Data for protein and acetylation events were collapsed to gene symbols of the relevant proteins, keeping the highest prized features as inputs for OI2. Due to the large number of differential phosphosites, these sites were treated as individual nodes in the OI2 network formulation and the required edges were added to the interactome (site-protein edges, cost: 0.25), splitting doubly phosphorylated sites into individual sites.

We ran OI2 with the selected disease-associated nodes and their prizes (see Table S6) as inputs on all combinations of the following parameters: Gs = (50k, 100k, 500k, 1M, 2.5M, 5M, 10M), Bs = (0.5, 1, 10), and Ws = (1, 3, 6, 10). We evaluated these networks to find a parameter set (W = 1, B = 1, G = 2.5M) that produced a network that was not dominated by hub-nodes, and had a reasonable balance of input nodes and those added by the algorithm for connectivity. Using this parameter set, we performed additional calculations to assess the robustness and specificity of nodes in the network. To determine the robustness of nodes, we added Gaussian noise (standard deviation: 0.05, mean: 0) to the edges in the interactome before each of 200 runs. We then calculated the robustness score for each node as the fraction of times a node appeared across the networks. To calculate specificity, we assigned our prizes to randomly selected, degree-matched nodes in the interactome before each of 200 runs. The specificity score was then calculated as the one minus the fraction of times a node appeared across these networks. (More details on and rationale around these scores can be found in (Tuncbag et al., 2016). For downstream analysis, we removed nodes from the OI2

output that were neither specific nor robust (threshold: 0.75). To link the network clusters to functional annotations we used hypergeometric tests as described above.

We used Cytoscape v3.3 (Shannon et al., 2003) for visualizations of our networks. Nodes received log2fold change-based colors based on the source of their prizes, and shapes according to the data type they represent (protein, genomic lesion, phosphosite, acetylated protein – see figure legend for details). Nodes in the network that did not receive a prize but were still included in the network received the log2-FC based color of the corresponding proteomic value if data was available (or were colored grey otherwise). For the final display items in Figure 7, we rearranged nodes primarily based on pathway associations rather than the clustering of the network used for the initial analysis.

Kinome Analysis—This analysis sought to identify the kinases potentially responsible for phosphorylating phosphopeptides of interest. We first selected phosphopeptides that were either specific to one of the clusters found in proteomic data or significantly different between a pair of clusters (FDR < 0.01). We then used two different methods to identify upstream kinases:

1. PhosphoSitePlus: Cell Signaling Technologies' PhosphoSitePlus database (Hornbeck et al., 2015) is a curated resource of experimentally determined peptide targets of specific kinases (downloaded February 2017). We searched this database for matches to phosphopeptides identified in our analyses.

2. Scansite: the Scansite platform (Obenauer et al., 2003) contains sequence specificity motifs for kinases that were derived from oriented peptide library screens. For this analysis, we used 15 amino acid long peptide sequences (phosphosite ± 7 amino acids) surrounding our phosphorylation events as input into Scansite (Scansite 3 Web Service) and used its most stringent setting (high stringency) to get only the best motif matches. For results to be reported with this setting, a motif's score for a peptide sequence needs to be in the top percentile of scores from an empirical score distribution that is based on all potentially matching sites in the vertebrate proteome.

For each of these two methods, we then collapsed all peptides matched to a kinase (median across matching peptides) and calculated the Spearman correlation of this 'peptide-profile' to the kinase's protein and/or phosphopeptide-levels (if available). If the correlation with the kinase's protein levels or any of its phosphopeptides was higher than 0.4 (or less than ⁻0.4 for phosphosites) it was labeled as correlating (or anticorrelating). Our peptide-kinase associations can be found in Table S5.

We checked DrugBank.ca (Law et al., 2014) for drugs that are known to target any of the kinases nominated by our analysis. Whenever this information was available, we also annotated whether any FDA approved products were listed for a drug. We provide a full table of all kinase-drug matches in Table S5.

To identify disease-related kinases that we could not associate with peptides (as in the analyses described above) due to a lack of supporting data (no kinase motifs or known sites),

we also extracted significantly differential kinases from all our proteomic data sets (FDR < 0.005). These kinases are indicated in Table S2.

We also used an R implementation (Wagih et al., 2016) of the motif-x algorithm (Chou and Schwartz, 2011) to discover motifs. We performed the analysis for 15 amino acid-long peptide sequence windows around serine (S), threonine (T), and tyrosine (Y) residues in our phospho-proteomics data sets. As foreground sets we used peptides up/down in only a single proteomic group of samples and ran the algorithm for all our groups at various significance thresholds with all S, T, Y-centered peptides in our data set as background. The most specific motifs found by this approach were "SP" and "TP".

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

REFERENCES

An J, Yang DY, Xu QZ, Zhang SM, Huo YY, Shang ZF, Wang Y, Wu DC, and Zhou PK (2008). DNA-dependent protein kinase catalytic subunit modulates the stability of c-Myc oncoprotein. Mol Cancer 7, 32. [PubMed: 18426604]

Arcella A, Carpinelli G, Battaglia G, D'Onofrio M, Santoro F, Ngomba RT, Bruno V, Casolini P, Giangaspero F, and Nicoletti F (2005). Pharmacological blockade of group II metabotropic glutamate receptors reduces the growth of glioma cells in vivo. Neuro Oncol 7, 236–245. [PubMed: 16053698]

Archer TC, Jin J, and Casey ES (2011). Interaction of Sox1, Sox2, Sox3 and Oct4 during primary neurogenesis. Dev Biol 350, 429–440. [PubMed: 21147085]

Archer TC, Mahoney EL, and Pomeroy SL (2017). Medulloblastoma: Molecular Classification-Based Personal Therapeutics. Neurotherapeutics 14, 265–273. [PubMed: 28386677]

Arnold HK, Zhang X, Daniel CJ, Tibbitts D, Escamilla-Powers J, Farrell A, Tokarz S, Morgan C, and Sears RC (2009). The Axin1 scaffold protein promotes formation of a degradation complex for cMyc. EMBO J 28, 500–512. [PubMed: 19131971]

Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, and Irizarry RA (2014). Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. Bioinformatics 30, 1363–1369. [PubMed: 24478339]

Barbie DA, Tamayo P, Boehm JS, Kim SY, Moody SE, Dunn IF, Schinzel AC, Sandy P, Meylan E, Scholl C, et al. (2009). Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. Nature 462, 108–112. [PubMed: 19847166]

Bigner DD, Bigner SH, Friedman HS, Oakes WJ, and Vogelstein B (1990). Amplification of the cmyc Gene in Human Medulloblastoma Cell Lines and Xenografts. Cancer Res.

Brown EJ, and Frazier WA (2001). Integrin-associated protein (CD47) and its ligands. Trends Cell Biol 11, 130–135. [PubMed: 11306274]

Cavalli FMG, Remke M, Rampasek L, Peacock J, Shih DJH, Luu B, Garzia L, Torchia J, Nor C, Morrissy AS, et al. (2017). Intertumoral Heterogeneity within Medulloblastoma Subgroups. Cancer Cell 31, 737–754 e6. [PubMed: 28609654]

Chen Y, and Olopade OI (2008). MYC in breast tumor progression. Expert Rev Anticancer Ther 8, 1689–1698. [PubMed: 18925859]

Cho Y-J, Tsherniak A, Tamayo P, Santagata S, Ligon A, Greulich H, Berhoukim R, Amani V, Goumnerova L, Eberhart CG, et al. (2011). Integrative genomic analysis of medulloblastoma identifies a molecular subgroup that drives poor clinical outcome. J. Clin. Oncol 29, 1424–1430. [PubMed: 21098324]

Chou MF, and Schwartz D (2011). Biological sequence motif discovery using motif-x. Curr Protoc Bioinforma. Chapter 13, Unit 13 15–24.

Cibulskis K, Lawrence MS, Carter SL, Sivachenko A, Jaffe D, Sougnez C, Gabriel S, Meyerson M, Lander ES, and Getz G (2013). Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. Nat Biotechnol 31, 213–219. [PubMed: 23396013]

Ciszewski WM, Tavecchio M, Dastych J, and Curtin NJ (2014). DNA-PK inhibition by NU7441 sensitizes breast cancer cells to ionizing radiation and doxorubicin. Breast Cancer Res. Treat

Costello M, Pugh TJ, Fennell TJ, Stewart C, Lichtenstein L, Meldrim JC, Fostel JL, Friedrich DC, Perrin D, Dionne D, et al. (2013). Discovery and characterization of artifactual mutations in deep coverage targeted capture sequencing data due to oxidative DNA damage during sample preparation. Nucleic Acids Res 41, e67. [PubMed: 23303777]

Cui F, Fan R, Chen Q, He Y, Song M, Shang Z, Zhang S, Zhu W, Cao J, Guan H, et al. (2015). The involvement of c-Myc in the DNA double-strand break repair via regulating radiation-induced phosphorylation of ATM and DNA-PKcs activity. Mol. Cell. Biochem

Davidson-Pilon C, Kalderstam J, Kuhn B, Fiore-Gartland A, Parij A, Stark K, Anton S, Besson L, and Gadgil H (2017). Lifelines survival analysis in Python. 0.11.1.

DeLuca DS, Levin JZ, Sivachenko A, Fennell T, Nazaire MD, Williams C, Reich M, Winckler W, and Getz G (2012). RNA-SeQC: RNA-seq metrics for quality control and process optimization. Bioinformatics 28, 1530–1532. [PubMed: 22539670]

DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, et al. (2011). A framework for variation discovery and genotyping using nextgeneration DNA sequencing data. Nat Genet 43, 491–498. [PubMed: 21478889]

Diaz RJ, Golbourn B, Faria C, Picard D, Shih D, Raynaud D, Leadly M, MacKenzie D, Bryant M, Bebenek M, et al. (2015). Mechanism of action and therapeutic efficacy of Aurora kinase B inhibition in MYC overexpressing medulloblastoma. Oncotarget.

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 29, 15–21. [PubMed: 23104886]

Edwards NJ, Oberti M, Thangudu RR, Cai S, McGarvey PB, Jacob S, Madhavan S, and Ketchum KA (2015). The CPTAC Data Portal: A Resource for Cancer Proteomics Research. J Proteome Res 14, 2707–2713. [PubMed: 25873244]

Farrell AS, and Sears RC (2014). MYC degradation. Cold Spring Harb. Perspect. Med

Gholamin S, Mitra SS, Feroze AH, Liu J, Kahn SA, Zhang M, Esparza R, Richard C, Ramaswamy V, Remke M, et al. (2017). Disrupting the CD47-SIRPalpha anti-phagocytic axis by a humanized antiCD47 antibody is an efficacious treatment for malignant pediatric brain tumors. Sci Transl Med 9.

Gibson P, Tong Y, Robinson G, Thompson MC, Currle DS, Eden C, Kranenburg TA, Hogg T, Poppleton H, Martin J, et al. (2010). Subtypes of medulloblastoma have distinct developmental origins. Nature 468, 1095–1099. [PubMed: 21150899]

Hanaford AR, Archer TC, Price A, Kahlert UD, Maciaczyk J, Nikkhah G, Kim JW, Ehrenberger T, Clemons PA, Dan ík V, et al. (2016). DiSCoVERing innovative therapies for rare tumors: Combining genetically accurate disease models with in silico analysis to identify novel therapeutic targets. Clin. Cancer Res 22.

Hornbeck PV, Zhang B, Murray B, Kornhauser JM, Latham V, and Skrzypek E (2015). PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. Nucleic Acids Res 43, D512–20. [PubMed: 25514926]

Horvath S, Mah V, Lu AT, Woo JS, Choi OW, Jasinska AJ, Riancho JA, Tung S, Coles NS, Braun J, et al. (2015). The cerebellum ages slowly according to the epigenetic clock. Aging (Albany NY) 7, 294–306. [PubMed: 26000617]

Hovestadt V, Jones DT, Picelli S, Wang W, Kool M, Northcott PA, Sultan M, Stachurski K, Ryzhova M, Warnatz HJ, et al. (2014). Decoding the regulatory landscape of medulloblastoma using DNA methylation sequencing. Nature 510, 537–541. [PubMed: 24847876]

Huang KL, Li S, Mertins P, Cao S, Gunawardena HP, Ruggles KV, Mani DR, Clauser KR, Tanioka M, Usary J, et al. (2017). Proteogenomic integration reveals therapeutic targets in breast cancer xenografts. Nat Commun 8, 14864. [PubMed: 28348404]

Jaiswal S, Jamieson CH, Pang WW, Park CY, Chao MP, Majeti R, Traver D, van Rooijen N, and Weissman IL (2009). CD47 is upregulated on circulating hematopoietic stem cells and leukemia cells to avoid phagocytosis. Cell 138, 271–285. [PubMed: 19632178]

Janghorban M, Farrell AS, Allen-Petersen BL, Pelz C, Daniel CJ, Oddo J, Langer EM, Christensen DJ, and Sears RC (2014). Targeting c-MYC by antagonizing PP2A inhibitors in breast cancer. Proc Natl Acad Sci U S A 111, 9157–9162. [PubMed: 24927563]

Kahn M (2014). Can we safely target the WNT pathway? Nat Rev Drug Discov 13, 513–532. [PubMed: 24981364]

Kanehisa M, Furumichi M, Tanabe M, Sato Y, and Morishima K (2017). KEGG: new perspectives on genomes, pathways, diseases and drugs. Nucleic Acids Res 45, D353–D361. [PubMed: 27899662]

Karolchik D, Hinrichs AS, Furey TS, Roskin KM, Sugnet CW, Haussler D, and Kent WJ (2004). The UCSC Table Browser data retrieval tool. Nucleic Acids Res 32, D493–6. [PubMed: 14681465]

Kim JW, Botvinnik OB, Abudayyeh O, Birger C, Rosenbluh J, Shrestha Y, Abazeed ME, Hammerman PS, DiCara D, Konieczkowski DJ, et al. (2016). Characterizing genomic alterations in cancer by complementary functional associations. Nat Biotechnol 34, 539–546. [PubMed: 27088724]

Kool M, Korshunov A, Remke M, Jones DT, Schlanstein M, Northcott PA, Cho YJ, Koster J, Schouten-van Meeteren A, van Vuurden D, et al. (2012). Molecular subgroups of medulloblastoma: an international meta-analysis of transcriptome, genetic aberrations, and clinical data of WNT, SHH, Group 3, and Group 4 medulloblastomas. Acta Neuropathol 123, 473–484. [PubMed: 22358457]

Kool M, Jones DT, Jager N, Northcott PA, Pugh TJ, Hovestadt V, Piro RM, Esparza LA, Markant SL, Remke M, et al. (2014). Genome sequencing of SHH medulloblastoma predicts genotyperelated response to smoothened inhibition. Cancer Cell 25, 393–405. [PubMed: 24651015]

Law V, Knox C, Djoumbou Y, Jewison T, Guo AC, Liu Y, Maciejewski A, Arndt D, Wilson M, Neveu V, et al. (2014). DrugBank 4.0: shedding new light on drug metabolism. Nucleic Acids Res 42, D1091–7. [PubMed: 24203711]

Lawrence RT, Perez EM, Hernandez D, Miller CP, Haas KM, Irie HY, Lee SI, Blau CA, and Villen J (2015). The proteomic landscape of triple-negative breast cancer. Cell Rep 11, 630–644. [PubMed: 25892236]

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 H, and Durbin R (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25, 1754–1760. [PubMed: 19451168]

Liberzon A (2014). A description of the Molecular Signatures Database (MSigDB) Web site. Methods Mol Biol 1150, 153–160. [PubMed: 24743996]

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

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

Louis DN, Ohgaki H, Wiestler OD, and Cavenee WK (2016b). WHO Classification of Tumours of the Central Nervous System (Geneva, Switzerland: WHO Press).

Lu AT, Hannon E, Levine ME, Hao K, Crimmins EM, Lunnon K, Kozlenkov A, Mill J, Dracheva S, and Horvath S (2016). Genetic variants near MLST8 and DHX57 affect the epigenetic age of the cerebellum. Nat Commun 7, 10561. [PubMed: 26830004]

Ma Y, Lu H, Tippin B, Goodman MF, Shimazaki N, Koiwai O, Hsieh CL, Schwarz K, and Lieber MR (2004). A biochemically defined system for mammalian nonhomologous DNA end joining. Mol Cell 16, 701–713. [PubMed: 15574326]

McCubrey JA, Steelman LS, Bertrand FE, Davis NM, Sokolosky M, Abrams SL, Montalto G, D'Assoro AB, Libra M, Nicoletti F, et al. (2014). GSK-3 as potential target for therapeutic intervention in cancer. Oncotarget 5, 2881–2911. [PubMed: 24931005]

Mermel CH, Schumacher SE, Hill B, Meyerson ML, Beroukhim R, and Getz G (2011). GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. Genome Biol 12, R41. [PubMed: 21527027]

Mertins P, Qiao JW, Patel J, Udeshi ND, Clauser KR, Mani DR, Burgess MW, Gillette MA, Jaffe JD, and Carr SA (2013). Integrated proteomic analysis of post-translational modifications by serial enrichment. Nat Methods 10, 634–637. [PubMed: 23749302]

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]

Morrish F, and Hockenbery D (2014). MYC and mitochondrial biogenesis. Cold Spring Harb Perspect Med 4.

Northcott PA, Hielscher T, Dubuc A, Mack S, Shih D, Remke M, Al-Halabi H, Albrecht S, Jabado N, Eberhart CG, et al. (2011). Pediatric and adult sonic hedgehog medulloblastomas are clinically and molecularly distinct. Acta Neuropathol 122, 231–240. [PubMed: 21681522]

Northcott PA, Jones DT, Kool M, Robinson GW, Gilbertson RJ, Cho YJ, Pomeroy SL, Korshunov A, Lichter P, Taylor MD, et al. (2012). Medulloblastomics: the end of the beginning. Nat Rev Cancer 12, 818–834. [PubMed: 23175120]

Northcott PA, Buchhalter I, Morrissy AS, Hovestadt V, Weischenfeldt J, Ehrenberger T, Gröbner S, Segura-Wang M, Zichner T, Rudneva VA, et al. (2017). The whole-genome landscape of medulloblastoma subtypes. Nature 547.

Obenauer JC, Cantley LC, and Yaffe MB (2003). Scansite 2.0: Proteome-wide prediction of cell signaling interactions using short sequence motifs. Nucleic Acids Res 31, 3635–3641. [PubMed: 12824383]

Pei Y, Liu KW, Wang J, Garancher A, Tao R, Esparza LA, Maier DL, Udaka YT, Murad N, Morrissy S, et al. (2016). HDAC and PI3K Antagonists Cooperate to Inhibit Growth of MYC-Driven Medulloblastoma. Cancer Cell 29, 311–323. [PubMed: 26977882]

Pfister S, Remke M, Benner A, Mendrzyk F, Toedt G, Felsberg J, Wittmann A, Devens F, Gerber NU, Joos S, et al. (2009). Outcome prediction in pediatric medulloblastoma based on DNA copynumber aberrations of chromosomes 6q and 17q and the MYC and MYCN loci. J Clin Oncol 27, 1627–1636. [PubMed: 19255330]

Pozo K, and Bibb JA (2016). The Emerging Role of Cdk5 in Cancer. Trends Cancer 2, 606–618. [PubMed: 27917404]

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]

Ramos AH, Lichtenstein L, Gupta M, Lawrence MS, Pugh TJ, Saksena G, Meyerson M, and Getz G (2015). Oncotator: cancer variant annotation tool. Hum Mutat 36, E2423–9. [PubMed: 25703262]

Rauniyar N, and Yates JR, 3rd (2014). Isobaric labeling-based relative quantification in shotgun proteomics. J Proteome Res 13, 5293–5309. [PubMed: 25337643]

Razick S, Magklaras G, and Donaldson IM (2008). iRefIndex: a consolidated protein interaction database with provenance. BMC Bioinformatics 9, 405. [PubMed: 18823568]

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]

Robinson GW, Orr BA, Wu G, Gururangan S, Lin T, Qaddoumi I, Packer RJ, Goldman S, Prados MD, Desjardins A, et al. (2015). Vismodegib Exerts Targeted Efficacy Against Recurrent Sonic Hedgehog-Subgroup Medulloblastoma: Results From Phase II Pediatric Brain Tumor Consortium Studies PBTC-025B and PBTC-032. J Clin Oncol 33, 2646–2654. [PubMed: 26169613]

Schwalbe EC, Lindsey JC, Nakjang S, Crosier S, Smith AJ, Hicks D, Rafiee G, Hill RM, Iliasova A, Stone T, et al. (2017). Novel molecular subgroups for clinical classification and outcome prediction in childhood medulloblastoma: a cohort study. Lancet Oncol 18, 958–971. [PubMed: 28545823]

Senbabaoglu Y, Michailidis G, and Li JZ (2014). Critical limitations of consensus clustering in class discovery. Sci Rep 4, 6207. [PubMed: 25158761]

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 Res 13, 2498–2504. [PubMed: 14597658]

Staal JA, Lau LS, Zhang H, Ingram WJ, Hallahan AR, Northcott PA, Pfister SM, Wechsler-Reya RJ, Rusert JM, Taylor MD, et al. (2015). Proteomic profiling of high risk medulloblastoma reveals functional biology. Oncotarget 6, 14584–14595. [PubMed: 25970789]

Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, et al. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A 102, 15545–15550. [PubMed: 16199517]

Sunada S, Kanai H, Lee Y, Yasuda T, Hirakawa H, Liu C, Fujimori A, Uesaka M, and Okayasu R (2016). Nontoxic concentration of DNA-PK inhibitor NU7441 radio-sensitizes lung tumor cells with little effect on double strand break repair. Cancer Sci.

Tamayo P, Cho Y-J, Tsherniak A, Greulich H, Ambrogio L, Schouten-van Meeteren N, Zhou T, Buxton A, Kool M, Meyerson M, et al. (2011). Predicting relapse in patients with medulloblastoma by integrating evidence from clinical and genomic features. J. Clin. Oncol 29, 1415–1423. [PubMed: 21357789]

Taylor MD, Northcott PA, Korshunov A, Remke M, Cho YJ, Clifford SC, Eberhart CG, Parsons DW, Rutkowski S, Gajjar A, et al. (2012). Molecular subgroups of medulloblastoma: the current consensus. Acta Neuropathol 123, 465–472. [PubMed: 22134537]

Tuncbag N, Gosline SJ, Kedaigle A, Soltis AR, Gitter A, and Fraenkel E (2016). Network-Based Interpretation of Diverse High-Throughput Datasets through the Omics Integrator Software Package. PLoS Comput Biol 12, e1004879. [PubMed: 27096930]

Wagih O, Sugiyama N, Ishihama Y, and Beltrao P (2016). Uncovering Phosphorylation-Based Specificities through Functional Interaction Networks. Mol Cell Proteomics 15, 236–245. [PubMed: 26572964]

Wang X, Cunningham M, Zhang X, Tokarz S, Laraway B, Troxell M, and Sears RC (2011). Phosphorylation regulates c-Myc's oncogenic activity in the mammary gland. Cancer Res 71, 925–936. [PubMed: 21266350]

Weeraratne SD, Amani V, Teider N, Pierre-Francois J, Winter D, Kye MJ, Sengupta S, Archer T, Remke M, Bai AH, et al. (2012). Pleiotropic effects of miR-183~96~182 converge to regulate cell survival, proliferation and migration in medulloblastoma . Acta Neuropathol. .

Wilkerson MD, and Hayes DN (2010). ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. Bioinformatics 26, 1572–1573. [PubMed: 20427518]

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, Meltzer P, and Davis S (2013). RCircos: an R package for Circos 2D track plots. BMC Bioinformatics 14, 244. [PubMed: 23937229]

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 X, Farrell AS, Daniel CJ, Arnold H, Scanlan C, Laraway BJ, Janghorban M, Lum L, Chen D, Troxell M, et al. (2012). Mechanistic insight into Myc stabilization in breast cancer involving aberrant Axin1 expression. Proc Natl Acad Sci U S A 109, 2790–2795. [PubMed: 21808024]

Highlights:

1. Deep proteomic profiling reveals mechanistic differences among medulloblastomas

2. Proteomic defined SHH subtypes do not differ in RNA expression

3. MYC phosphorylation events define a higher risk subset of Group 3 patients

4. Inhibiting PRKDC may sensitize MYC-activated medulloblastoma tumors to radiation
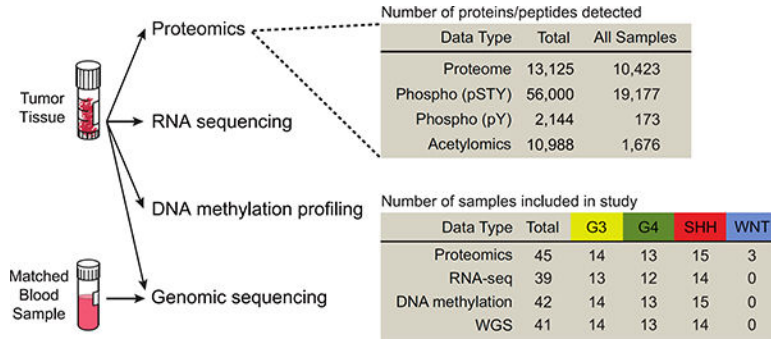
Figure 1: Summary of data types included in this study, depth of proteomic data types, and cohort composition.

The extent of the data from proteomics, including post-translational modifications, is summarized at the top. pSTY: phosphorylation on serine, threonine, or tyrosine detected after immobilized metal affinity chromatography; pY: phosphorylated tyrosine detected after antibody purification; Total: the total number of features identified; All Samples: the number of features measured across all samples, i.e. without any missing values. The number of samples covered by each data type and their split by subgroups are summarized at the bottom. G3: Group 3; G4: Group 4; WGS: whole genome sequencing. Proteomics includes proteome, pSTY, pY, and acetylomics data sets. See also Table S1.
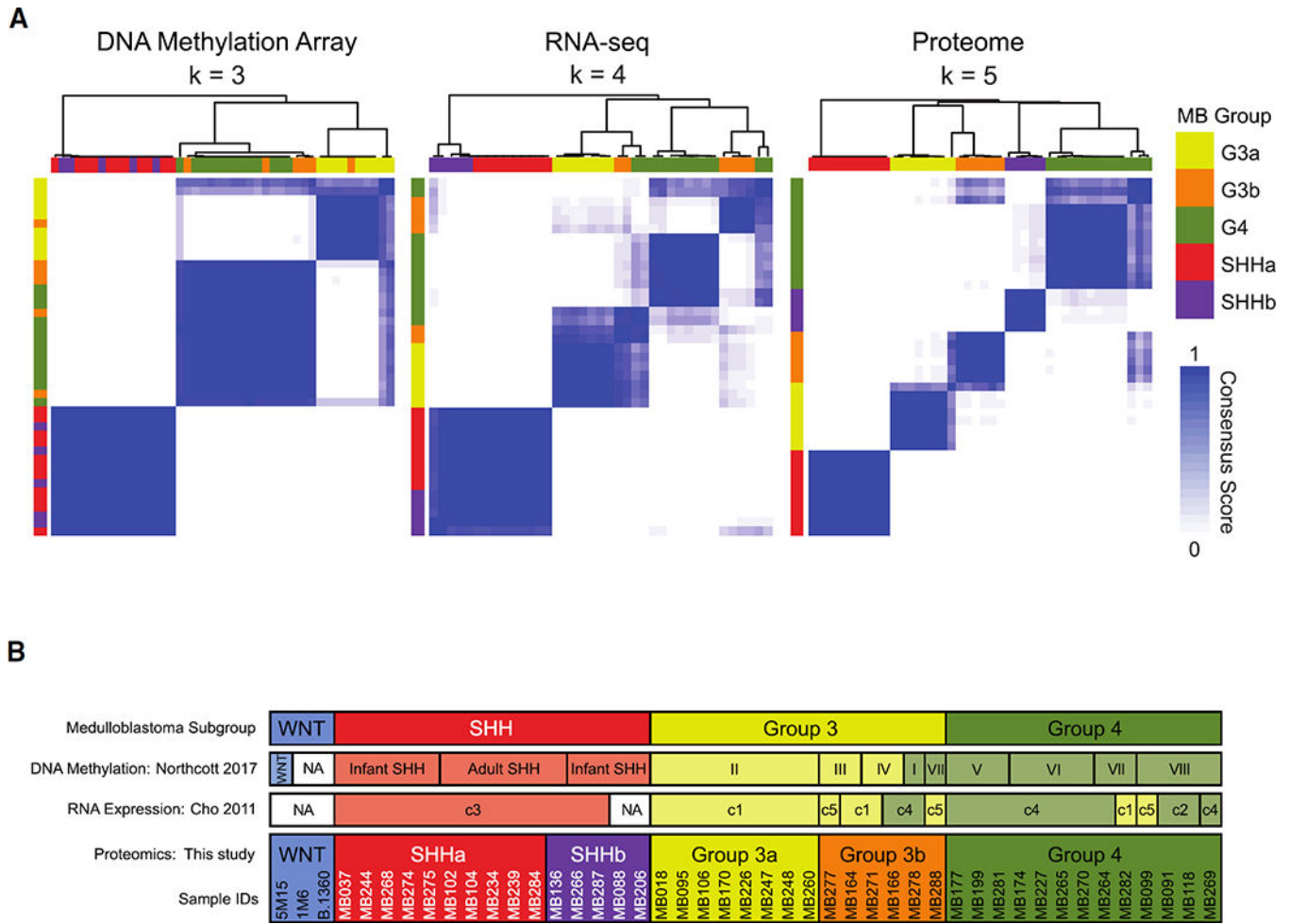
Figure 2: Comparison of clustering results.

(**A**) The optimal clustering of DNA methylation data, RNAseq, and proteome, as determined using Pearson correlation as distance metric. *k*, number of clusters. Consensus scores are indicated using a color scale from white (samples never cluster together) to blue (samples always cluster together). (**B**) Comparison of the assignment of samples using the four "consensus subgroups" (Taylor et al., 2012), the DNA Methylation-based subtype-calls assigned in Northcott *et al.* (2017), which included most samples used in our study, and RNA expression assignments based on application of the classifier described in Cho *et al.* (2011). NA: no assignment available. See also Figures S1-S3 and Tables S1 and S2.
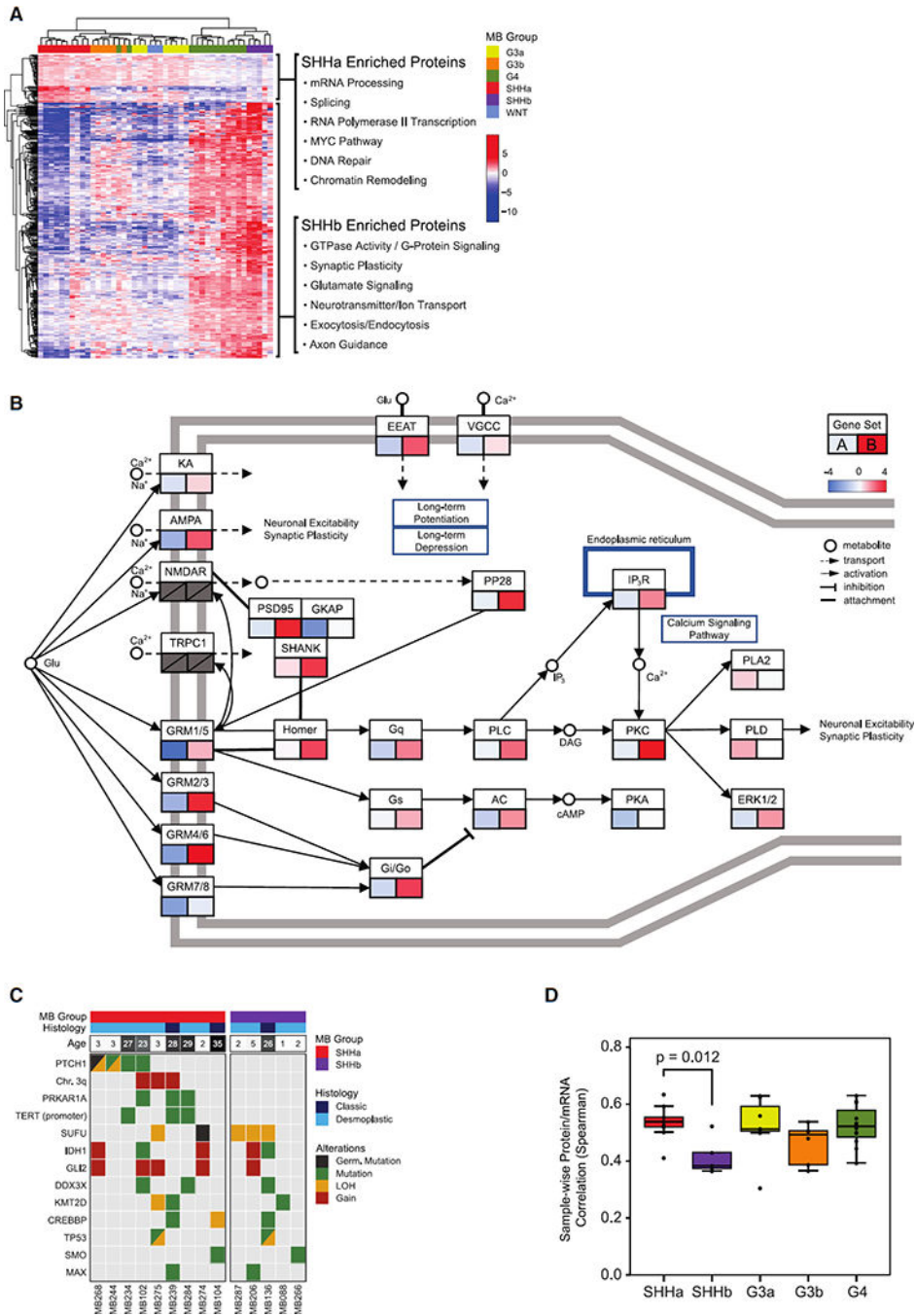
Figure 3: Molecular and functional differences between SHHa and SHHb.
(**A**) Hierarchical clustering of 510 proteomic features that differ significantly between SHHa and SHHb (FDR < 0.005; ANOVA). Summaries of gene sets that significantly differ (FDR < 0.01; hypergeometric overlap tests) between SHHa and SHHb are shown alongside the heatmap. (**B**) A schematic of the pathway for the glutamatergic synapse. Each box represents a gene or gene set taken directly from the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway. The left and right boxes summarize SHHa and SHHb protein levels, respectively. Grey boxes indicate a lack of proteomic data for the gene product. Small

white circles represent metabolites, solid directed arrows activation, and lines ending with a bar inhibition. Dotted lines indicate transport of metabolites. Solid lines and connected boxes represent physical interactions in the cell. Figure adapted from KEGG pathway hsa04724 (Kanehisa et al., 2017). (**C**) Clinical annotations of and recurrent genomic alterations in SHHa and SHHb samples. (**D**) Box-and-whisker plots of Spearman correlations between the mRNA and protein levels for approximately 8,700 genes for each tumor sample. Boxes range from first to third quartile in each group, with line indicating the group-median. The whiskers extend to the lowest and highest data points within 1.5 times the interquartile range from the box (p value shown from Mann Whitney U test). See also Figure S4 and Tables S3 and S4.
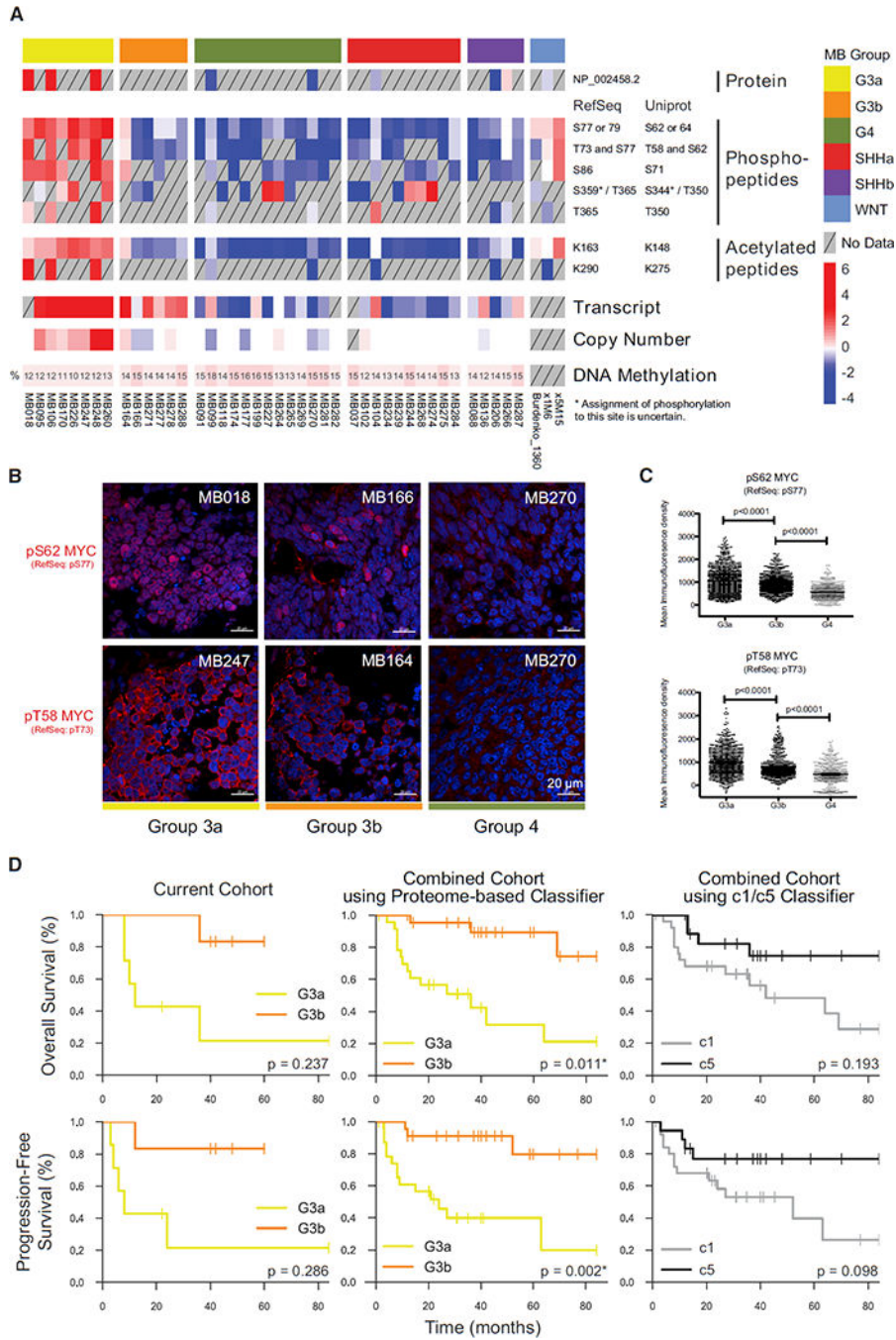
Figure 4: G3a medulloblastomas have post-translational modifications of MYC protein that strongly correlate with MYC downstream activity and patient outcome.
(**A**) Heatmap of multi-omic data for MYC, showing protein levels, post-translationally modified peptides, mRNA expression, DNA copy number, and DNA methylation. Scale bar is of log2 values for all data sets except DNA methylation, which ranges from 0 to 1 on a linear scale. Methylation values, converted to percentages, are superimposed on the heatmap. The phosphorylation locations are provided both based on RefSeq (as used in our data sets) and Uniprot (frequently used in the literature). S77 or 79: phosphorylated at either

S77 or S79 but not at T73; S359* / T365: phosphorylated at T365 as well as a proximal site that could not be uniquely determined. (**B**) Representative confocal microscopy images showing pS62 and pT58 MYC in G3a, G3b, and Group 4 tumors. (**C**) Scatter bar charts showing mean immunofluorescence densities quantified across 50–100 nuclei in 3–4 fields of view captured at 40x for MYC pS62 and pT58. Error bars represent standard error of means (SEM) and significance reported is from two-tailed unpaired t-tests. (**D**) Kaplan-Meier plots of overall survival and progression-free survival of Group 3 medulloblastoma from the current cohort (left) and the current cohort combined with Group 3 samples from Cho *et al.* (2011). The survival analysis on the combined cohort was performed using the classifier defined by the proteomic clusters G3a and G3b ("Proteome-based Classifier" in the center) and the original classifier from Cho *et al.* (2011) (c1/c5, right, gray scale). In the combined cohort with proteomebased group assignments (G3a/G3b), both outcome predictions reached statistical significance (p values as shown; Log-rank test). See also Figure S5.
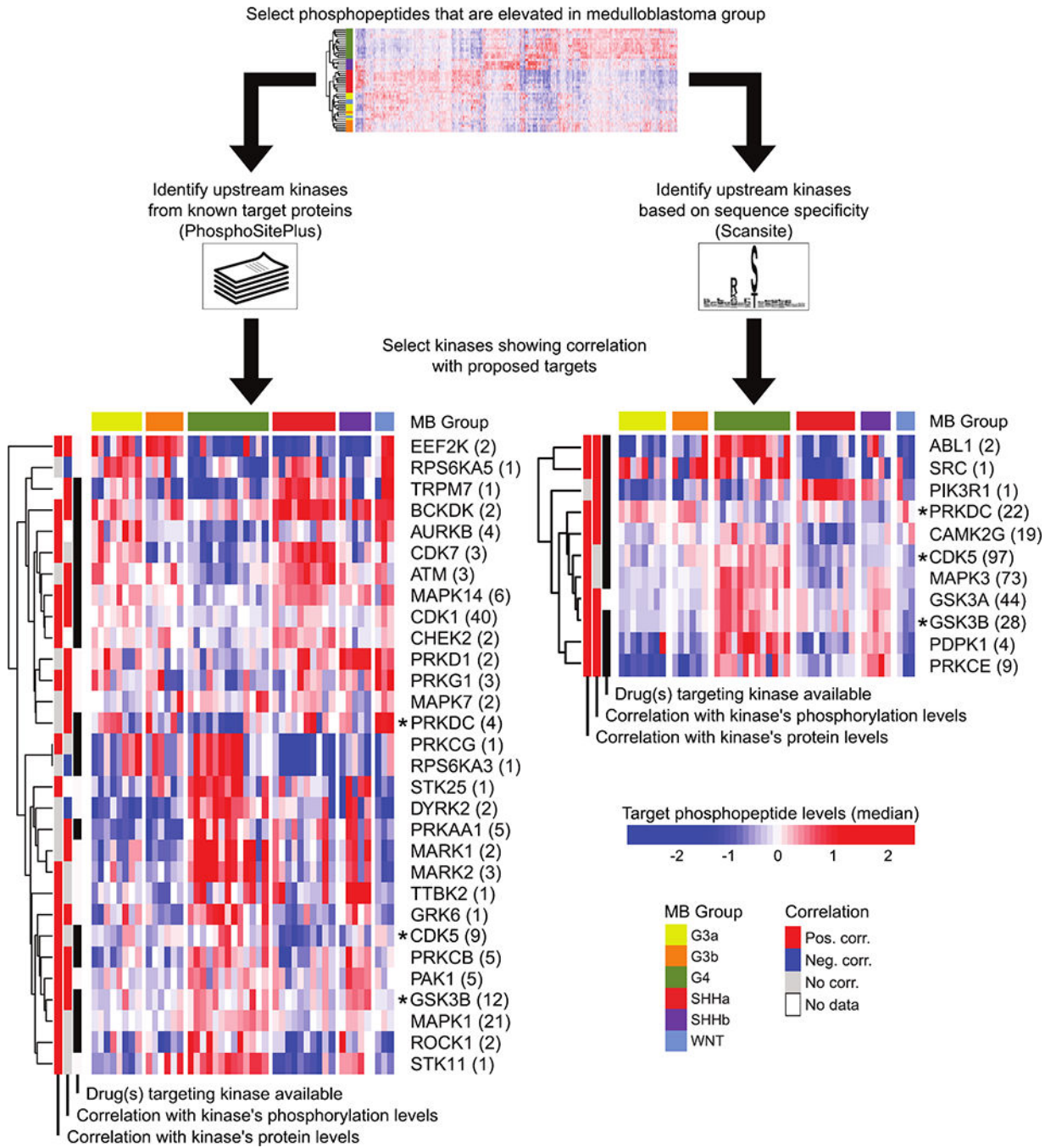
Figure 5: Medulloblastoma subtypes differ in kinase regulation and substrate levels.
Two different methods used to analyze the kinome are shown. On the left, upstream kinases are predicted from differential phosphopeptides between subgroups using PhosphoSitePlus database (Hornbeck et al., 2015). On the right, kinases are predicted from scoring differential phosphopeptides using sequence specificity motifs from Scansite (Obenauer et al., 2003). Heatmaps show the median levels of peptides matched to an upstream kinase, with the number of peptides matching each kinase shown in parentheses. Kinases found by both methods are annotated with an asterisk. Bars on the side of the heatmaps indicate
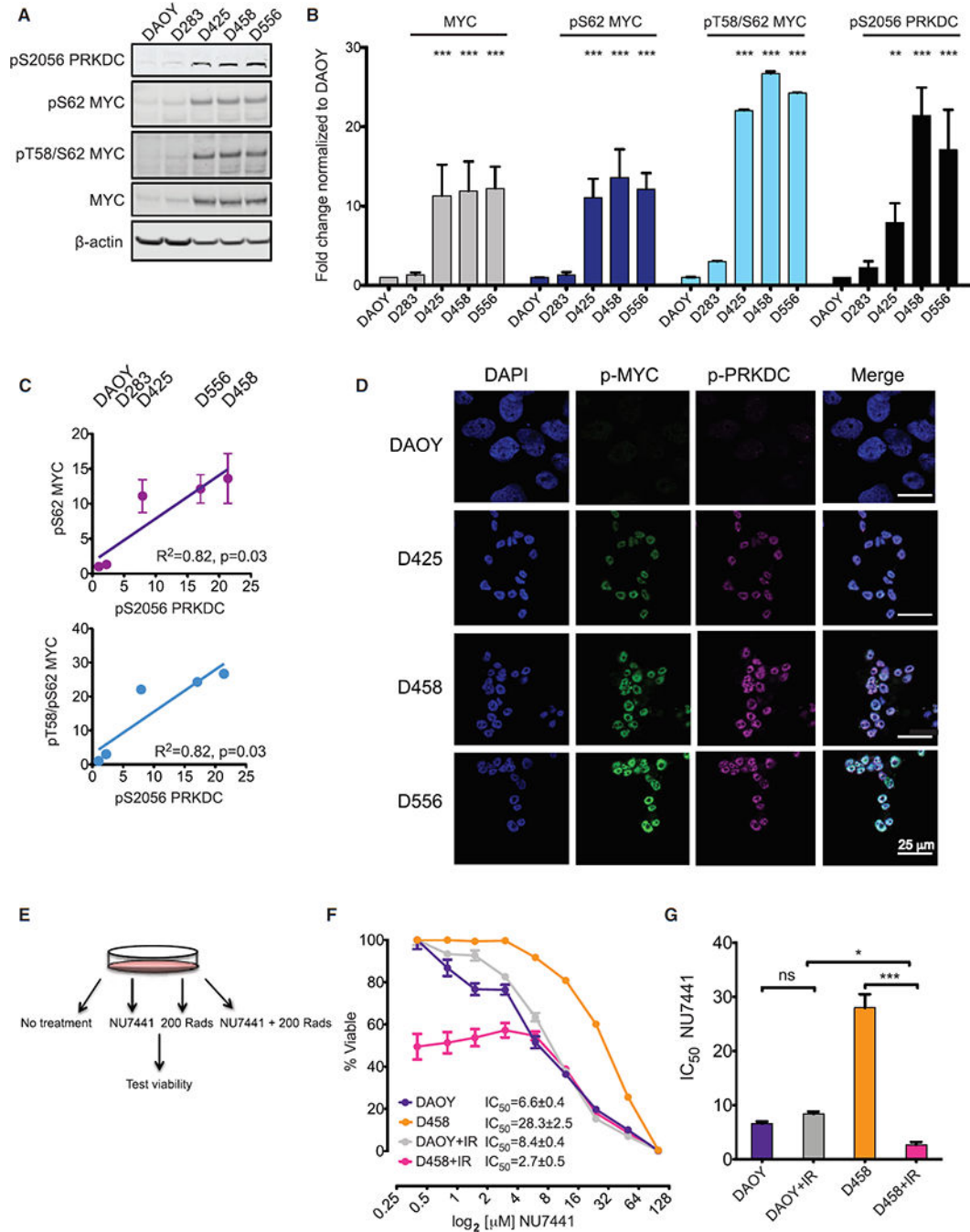
whether target peptides correlate with protein or phosphorylation levels of upstream kinases; and if DrugBank (Law et al., 2014) lists any drugs targeting the kinases. See also STAR Methods and Table S5.

Figure 6: MYC status correlates with PRKDC phosphorylation, and predicts increased sensitivity to PRKDC inhibition with irradiation.

(**A**) Representative Western blots of medulloblastoma cell lines performed in at least triplicate. Antibody against pT58/S62 MYC detects either site or both sites. (**B**) Quantification of Western blots represented as fold change compared to expression in DAOY cells. All proteins are normalized to β-actin. Significance was determined via one-way ANOVA with a Dunnett multiple comparison test to compare the normalized signal for each antibody across the five cells lines; **p < 0.001, ***p < 0.0001. Error bars represent mean

normalized signal ± SEM. (**C**) Correlations between normalized means for specified antibodies determined by calculating a Pearson correlation coefficient. Error bars indicate ± SEM, and are depicted but are not visible for some data points because of scale. (**D**) Representative confocal images of indicated cell lines showing pS62 MYC and pS2056 PRKDC. (**E**) Experimental design of dose-response curve of PRKDC inhibitor NU7441 for 18 hours prior to irradiation. (**F**) Viability assay of medulloblastoma cell lines treated as indicated. Plotted is the mean of 6 biological replicates ± SEM for each dose; note, small error bars for D458 are depicted but obfuscated by trend lines. (**G**) Histogram of mean IC$_{50}$ values ± SEM for DAOY vs D458 treated with NU7441 ± irradiation. *p < 0.01; ***p < 0.0001; ns = not significant; IR = irradiation.
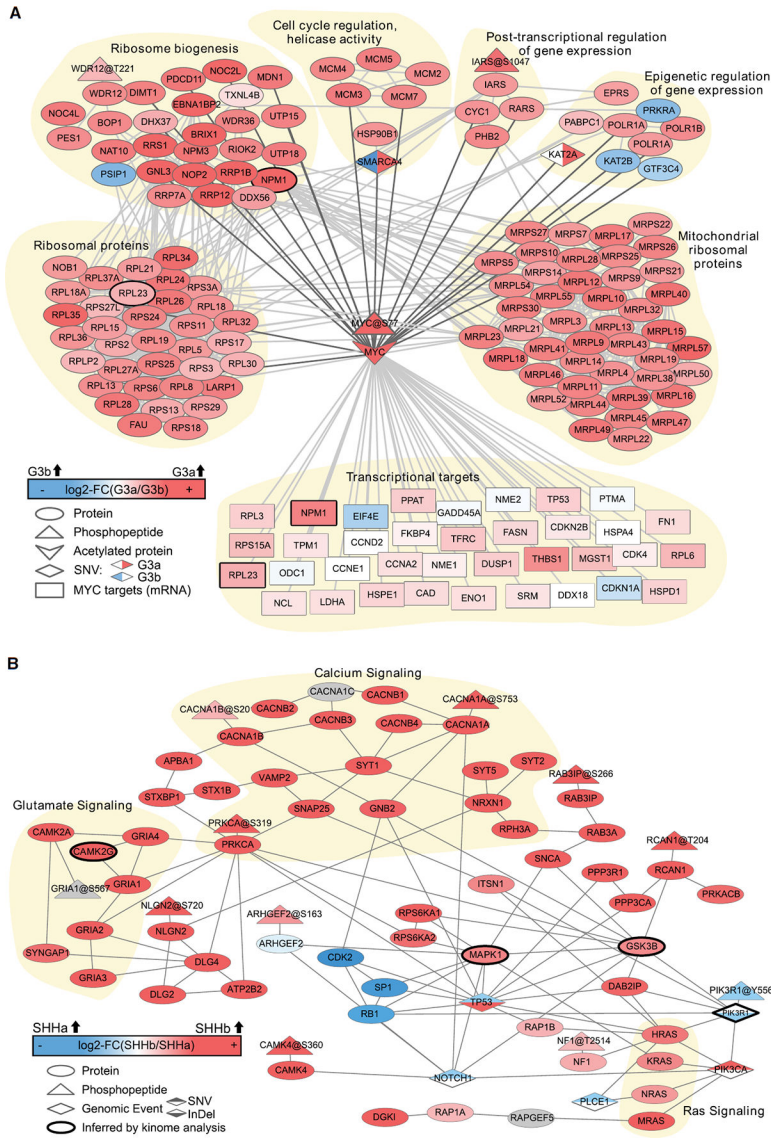
Figure 7: Network methods relying on known protein-protein interactions identify pathways relating to SHHb and G3a tumors.

Omics Integrator output showing network views of proteins, posttranslational modifications, and genomic alterations associated with G3a (**A**) and SHHb (**B**). Node shapes indicate data type and colors indicate log2-based fold change between groups as described in the legends. Phosphopeptides are labeled with their phosphorylation sites (based on RefSeq) after the '@' symbol. Nodes associated with selected pathways are highlighted with yellow background. Grey nodes were added by Omics Integrator and have no associated proteomic data for our samples. (**A**) mRNA levels of transcriptional targets of MYC are shown at the bottom. Thick borders highlight proteins that are also shown as direct transcriptional targets of MYC. SNVs, indicated by diamonds, are color-coded to show which subtype the genomic alteration was seen in: G3a, red; G3b, blue. (**B**) Kinases that were also found by our independent Kinome Analysis (Figure 5) are highlighted with thick borders. The color of genomic lesions (diamonds) indicates the subtype in which they occur: SHHb, red; SHHa,

blue. The location of color in a diamond indicates the type of genomic lesion: upper triangle, SNV; lower triangle, indel.

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Phospho-Tyrosine (P-Tyr-1000)	Cell Signaling Technology	8954; RRID:AB_2687925
PTMScan Acetyl-Lysine Motif [Ac-K] Kit #13416	Cell Signaling Technology	https://www.cellsignal.com/products/proteomic-analysis-products/acetyl-lysine-motif-ac-k-kit/13416
Rabbit polyclonal c-Myc pS62 specific antibody 1:25	Zhang *et al.* (2012)	
Rabbit polyclonal c-Myc pT58 antibody 1:50	Applied Biological Materials	Y011034; RRID:AB_895681
Anti-DNA PKcs (phospho S2056) antibody - ChIP Grade (ab18192) 1:500	Abcam	ab18192 RRID:AB_869495
p-c-Myc S62/T58 Antibody (C-3): sc-377551 1:500	Santa Cruz	sc-377551
β-Actin (8H10D10) Mouse mAb 1:2000	Cell Signaling Technology	3700S; RRID:AB_2242334
Phospho-c-Myc (Ser62) (E1J4K) Rabbit mAb 1:500	Cell Signaling Technology	13748S
c-Myc (D84C12) Rabbit mAb 1:1000	Cell Signaling Technology	5605S RRID:AB_1903938
Alexafluor 488 Goat anti Mouse	Thermo	A11029 RRID:AB_2534088
Alexafluor 647 Goat anti Rabbit	Thermo	A21244 RRID:AB_2535812
IRDye 800 CW Goat anti Rabbit	Licor	925–32211 RRID:AB_2651127
IRDye 680RD Goat anti Mouse	Licor	925–98070 RRID:AB_10956588
Reagent		
Licor Odyssey Blocking Buffer	Licor	927–50000
NuPAGE™ Novex™ 4–12% Bis-Tris Midi Protein Gels, 20-well	Thermo	WG1402BOX
iBlot2 PVDF Stacks	Thermo	IB24001
Bolt MES SDS running buffer	Thermo	B000202
Pierce BCA Protein Assay Kit	Thermo	23225
Benzonase	Santa Cruz	sc-202391
Phosphatase inhibitors	Thermo	88667
Protease inhibitors	Thermo	88266
NU7441	Selleck Chemicals	S2638
CellTiter-Glo Luminescent Cell Viability Assay	Promega	G7571
Biological Samples		
45 primary medulloblastoma samples	Northcott *et al.* (2017)	
Deposited Data		
WGS, RNA-seq, Affy expression, and DNA	Northcott *et al.* (2017)	http://www.ebi.ac.uk/ega/studies/EGAS00001001953

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Methylation data: EGAS00001001953		
Proteomics Data	http://massive.ucsd.edu	MSV000082644
Software and Algorithms		
OmicsIntegrator 2 (v0.2.24)	Tuncbag *et al.* (2016)	https://github.com/fraenkel-lab/OmicsIntegrator2
Cytoscape v3.3.0	Shannon *et al.* (2003)	RRID:SCR_003032
ConsensusClusterPlus Bioconductor library (v1.38.0)	Wilkerson and Hayes (2010)	http://bioconductor.org/packages/release/bioc/html/ConsensusClusterPlus.html
The R project (R v3.3.2)		RRID:SCR_001905
Minfi Bioconductor library (v1.16.1)	Aryee *et al.* (2014)	RRID:SCR_012830
Scansite Webservice	Obenauer, Cantley and Yaffe (2003)	RRID:SCR_001905
Illumina's 450k array probe annotation Bioconductor library (v0.2.1)		http://bioconductor.org/packages/release/data/annotation/html/IlluminaHumanMethylation450kanno
RMotifX	Wagih *et al.* (2016)	https://github.com/omarwagih/rmotifx
Rtsne		https://github.com/jkrijthe/Rtsne
Picard tools	http://broadinstitute.github.io/picard/	RRID:SCR_006525
STAR	Dobin *et al.* (2013)	https://github.com/alexdobin/STAR
RSEM	Li and Dewey (2011)	RRID:SCR_013027
RNASeqQC	DeLuca *et al.* (2012)	RRID:SCR_005120
BWAmem	Li and Durbin (2009)	RRID:SCR_010910
GATK	DePristo *et al.* (2011)	RRID:SCR_001876
MuTect2	Cibulskis *et al.* (2013)	RRID:SCR_000559
Oncotator	Ramos *et al.* (2015)	RRID:SCR_005183
D-ToxoG	Costello *et al.* (2013)	http://archive.broadinstitute.org/cancer/cga/dtoxog
GISTIC2.0	Mermel *et al.* (2011)	RRID:SCR_000151
ssGSEA	Barbie *et al.* (2009)	RRID:SCR_003199
Spectrum Mill		
RCircos	Zhang, Meltzer and Davis (2013)	https://cran.r-project.org/web/packages/RCircos/
PreprocessCore (v1.36)		http://bioconductor.org/packages/preprocessCore/
OpenLab 5.5	PerkinElmer Inc.	RRID:SCR_012158
ZEN Digital Imaging for Light Microscopy	Carl Zeiss	RRID:SCR_013672
Image Studio Lite for Western Blots	Licor	RRID:SCR_014211
GraphPad Prism		RRID:SCR_002798
Lifelines		https://github.com/CamDavidsonPilon/lifelines
Image J		RRID:SCR_003070
Kaleido 1.2	PerkinElmer Inc.	
Other		
PhosphoSitePlus (downloaded March 2017)	Hornbeck *et al.* (2015)	RRID:SCR_001837

REAGENT or RESOURCE	SOURCE	IDENTIFIER
DrugBank.ca (downloaded Oct 2017)	Law *et al.* (2014)	RRID:SCR_002700
MSigDb (v6.0)	Liberzon *et al.* (2015)	RRID:SCR_003199
iRefIndex (v14.0)	Razick, Magklaras and Donaldson (2008)	RRID:SCR_002085
UCSC RefSeq (downloaded Sept 14, 2016)	Karolchik *et al.* (2004)	https://genome.ucsc.edu/cgi-bin/hgTables
Odyssey CLx Infrared Imaging System	Licor	RRID:SCR_014579
EnSight Multimode Plate Reader	PerkinElmer Inc.	HH34000000