

Comprehensive Proteogenomic Characterization Reveals Clinically Relevant Molecular Subtypes Associated with Medulloblastoma Progression

Jong Bae Park

jbp700922@gmail.com

National Cancer Center

Seong-min Park

National Cancer Center

Kyunh-Hee Kim

National Cancer Center

Jong Hyuk Yoon

Korea Brain Research Institute <https://orcid.org/0000-0002-4090-3832>

Fulvio D'Angelo

University of Miami Miller School of Medicine

Seung Ah Choi

Seoul National University Hospital

Chan Il Kim

National Cancer Center

Harim Koo

Korea University College of Medicine

Seung Min Park

National Cancer Center

Hyondeog Kim

National Cancer Center

Sreeja Raj Sundara

National Cancer Center

Sung Soo Kim

National Cancer Center

Ae Kyung Park

Jeonbuk National University

Eun Jung Koh

Seoul National University Children's Hospital

Seong-Ik Kim

Seoul National University College of Medicine

Yu-Mi Shim

Seoul National University College of Medicine

Kwang Hoon Lee

Seoul National University Hospital

Ji Hoon Phi

Seoul National University Hospital

Yeon Suk Jo

Korea Brain Research Institute

Do-Hyun Nam

Sungkyunkwan University

Daehee Hwang

Seoul National University <https://orcid.org/0000-0002-7553-0044>

Do Young Hyeon

Seoul National University

Sunghyun Huh

Seoul National University

Hyung Joon Kwon

National Cancer Center

Seok Jun Ha

National Cancer Center

Sanha Park

National Cancer Center

Hyeji Shin

National Cancer Center

Jeong Taik Kwon

Chung-Ang University Hospital

Heon Yoo

National Cancer Center

Ho-Shin Gwak

National Cancer Center

Michael Taylor

Texas Children's Hospital <https://orcid.org/0000-0001-7009-3466>

Jason Sa

Korea University <https://orcid.org/0000-0002-3251-5004>

Youngwook Kim

National Cancer Center, Korea

Antonio Iavarone

Sylvester Comprehensive Cancer Center, University of Miami Miller School of Medicine

<https://orcid.org/0000-0002-0683-4634>

Sung-Hye Park

Seoul National University College of Medicine <https://orcid.org/0000-0002-8681-1597>

Seung-Ki Kim

Seoul National Univ.

Eric Kim

Seoul National University College of Medicine

Article

Keywords:

Posted Date: March 14th, 2025

DOI: <https://doi.org/10.21203/rs.3.rs-5954933/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: There is **NO** Competing Interest.

Abstract

Current treatment strategies for medulloblastoma remain ineffective due to extensive tumor heterogeneity. In this study, we performed integrated multi-omic characterization to improve the conventional molecular classification of medulloblastoma, leading to the identification of seven refined distinct subtypes. The SHH group was reclassified into two subgroups, SHHa and SHH β , while group 4 was divided into three subgroups, G4 α , G4 β , and G4 γ . SHH and Group 4 subtypes exhibit two distinct neuronal differentiation trajectories: granular neuron (GN) and unipolar brush cell (UBC) differentiation (SHH β and G4 γ , respectively), both of which associated with more favorable clinical outcome. Furthermore, we uncovered unique proteomic and kinomic properties that conferred increased treatment vulnerabilities to targeted therapeutic interventions against each of the three medulloblastoma subtypes associated with poor clinical outcome. We demonstrated the therapeutic potential of exploiting these vulnerabilities by utilizing a proteasome inhibitor and subtype-specific agents, including CDK1/2, PARP, CLK1, and MET inhibitors. Mechanistic insights were further elucidated through in-depth proteome analyses. In conclusion, our study qualifies the use of proteomic signatures and activation of neuronal differentiation trajectories to tailor selective therapeutic opportunities for distinct subgroups of medulloblastoma patients.

Main

Medulloblastoma is a representative malignant pediatric brain cancer for which the standard-of-care after surgery still involves conventional chemotherapy and radiotherapy, with significant toxicity and long-term morbidities ¹. Furthermore, medulloblastoma patients frequently experience tumor recurrence toward mostly incurable stages of the disease; according to previous studies, the overall 3-year survival after relapse is 18% and the 5-year survival only 6% ^{2,3}. Thus, there is an unmet requirement for more personalized therapeutic approaches in patients with medulloblastoma.

Various omics studies have been performed with the aim of improving the molecular diagnosis of medulloblastoma. Four distinct molecular subgroups of medulloblastoma have been defined based on genomic, transcriptomic and methylomic features (Wingless and Int-1 (WNT), Sonic hedgehog (SHH), Group 3 and Group 4), which are now accepted as the international consensus classification of medulloblastoma ⁴. The WNT and SHH subgroups are characterized by genomic mutations that activate the genetic drivers of respective pathways, including CTNNB1, PTCH1 and TP53 mutations, whereas Groups 3 and 4 lack significant mutations and are instead defined by unique transcriptomic signatures ⁵. By applying methylome analysis, the four subgroups were further subdivided into multiple subtypes ⁶. It was suggested that, unlike the other pediatric tumors that exhibit variable levels of immune infiltration, medulloblastoma is an immune cold tumor. ⁷. Recently, a single-cell genomics study revealed that Group 3 and Group 4 tumors show a developmental trajectory from primitive progenitor-like to more mature neuronal-like cells ⁸. Overall, multiomics studies have identified subtype markers and revealed different aspects of medulloblastoma.

Although the molecular hallmarks of medulloblastoma at diagnosis opened new possibility for the use of integrated multiomics data towards precision medicine applications, these studies failed to exhibit favorable impact for medulloblastoma treatment. In this study, we explored the molecular profiles of medulloblastoma specimens, which included matched longitudinal samples, by using five different multiomics platforms. By performing integrated proteogenomic analyses, we identified new subtypes, investigated the difference in progression patterns among the subtypes, and identified progression-related biomarkers and subtype specific therapeutic targets. Our findings provide the refined molecular background necessary to accelerate the development of precision therapy for patients with medulloblastoma.

Results

Multiomics analysis stratifies medulloblastoma patients into seven subtypes

We conducted multi-omics profiling of 123 medulloblastomas, including primary and recurrent tumors, from 102 patients (Extended Data Fig. 1). Our analysis involved five omics platforms: genomics, transcriptomics, global proteomics, phosphoproteomics, and methylomics (Extended Data Fig. 1). Using liquid chromatography and mass spectrometry (LC-MS/MS) and tandem mass tag (TMT) labeling, we generated global proteomic and phosphoproteomic data from 140 samples across 116 tumor tissues. The global proteomic analysis identified and quantified 12,963 proteins in at least one tumor sample and 8,409 unique proteins were detected across all samples. The phosphoproteomic analysis identified 47,233 unique phosphosites, where 8,094 proteins were detected in at least one sample and 3,763 phosphosites and 3,592 proteins across all samples. After filtering proteins or phosphosites that contained more than 30% missing values, the final output resulted in 10,124 proteins and 9,992 phosphosites for global proteome and phosphoproteome, respectively. Additionally, we generated genomic, methylomic, and transcriptomic data for 109, 106, and 112 tumor tissues, respectively, alongside genomic data from 80 matched blood. Two patients were excluded due to histological reclassification as astroblastoma. The comprehensive summary of the sample and data status is presented in Extended Data Fig. 1.

Previous studies defined four molecular subgroups of medulloblastoma, WNT, SHH, Group 3, and Group 4, based on genomic, transcriptomic, and methylomic data^{4,5}. We assessed whether our multi-omics data reproduced the conventional subgroups with an extended dataset. Unsupervised clustering methods, including hierarchical and non-negative matrix factorization (NMF) from transcriptomic and methylomic data from primary tumors, consistently identified the reported four subgroups (Extended Data Figs. 2 and 3). This robust reproducibility across different clustering methods confirmed that our cohort represents the conventional four molecular subgroups of medulloblastoma.

To identify a more granular and clinically relevant molecular resolution of medulloblastoma classification, we applied an integrative approach that combined the multi-omics data from primary

tumors (Fig. 1a). Our proteogenomics-based clustering revealed a pattern similar but with significant differences from the conventional molecular classifications (Extended Data Fig. 3 and Fig. 1). Within the SHH group, we identified two distinct subtypes, including SHH α and SHH β , with SHH β demonstrating high similarity to Group 4 (Extended Data Figs. 2 and 3). Due to the limited number of samples, further subclassification of Group 3 was not feasible. Conversely, we uncovered distinct variability within the Group 4 tumors, identifying three dominant subtypes, including G4 α , G4 β , and G4 γ . To validate our proteogenomic classification, we compared our results using a public proteogenomic dataset⁹ (Extended Data Fig. 4). As a result, we found significant correlations between the global and phosphoproteome datasets, confirming the consistency and robustness of our newly defined subtypes (Extended Data Fig. 4).

Additionally, we performed whole-exome and whole-genome sequencing to identify essential genomic aberrations that are unique to each subtype (Fig. 1a). Consistent with prior studies, we discovered subtype-specific genetic alterations, including *CTNNB1* and *PTCH1* mutations in WNT and SHH subtypes, respectively^{5,10}. To validate the expression patterns of representative RNA markers of the conventional subgroups, we analyzed the transcriptome profiles of the newly defined subtypes (Fig. 1a and Extended Data Fig. 5)¹¹. We also explored the functional activities underlying each subtype (Fig. 1a and Extended Data Fig. 6). Notably, the G4 α subtype was enriched with proteins associated with epithelial-mesenchymal transition (EMT) and extracellular matrix (ECM) reorganization functions, while G4 β and G4 γ tumors showed activation of neuron-related proteins. The SHH β tumors demonstrated enrichments of neuronal differentiation functions, including synaptic signaling and axon guidance, closely resembling the functional properties of G4 γ tumors. Conversely, the SHH α subtype was enriched with cell cycle-related functions. Collectively, our findings highlight the identification of novel medulloblastoma subtypes that constitute unique functional properties, including cell cycle regulation, neuronal function, EMT, and ECM organization. A detailed list of significant marker genes and proteins is included in Extended Data Fig. 6.

Next, we investigated the proteomic profiles of each subtype based on a protein-protein interaction network. Proteins associated with subtype-specific functions, such as neuronal system and EMT, marked distinct clusters within the network (Fig. 1b). Overlaying the abundance of subtype-specific proteins onto the network revealed a clear pattern where uniquely enriched proteins were localized within functionally relevant clusters. For example, G4 α -specific proteins were predominantly concentrated in EMT and coagulation cluster regions, whereas G4 γ subtype proteins were concentrated within the neuronal system in the network (Fig. 1b). We further validated such findings using public proteogenomic datasets, which showed similar functional clustering patterns (Extended Data Fig. 7). These results suggest that subtype-specific proteins undergo distinct functional modulation within the protein-protein interaction network.

We tracked the re-clustering of samples across different molecular layers (Fig. 1c and Extended Data Fig. 3). While the four conventional subgroups were consistently maintained in the methylome and transcriptome clusters, proteome-based classification identified seven distinct multi-omic clusters.

Survival analysis based on progression-free survival (PFS) rate demonstrated that the WNT, SHH β , and G4 γ subtypes were associated with favorable clinical outcomes, whereas other subtypes invariably showed worse survival probabilities ($p = 0.046$, Fig. 1d). Overall, these findings suggest that the newly defined proteogenomic-driven subtypes mark clinical relevance in medulloblastoma progression.

Medulloblastoma subtypes are associated with neuronal differentiation

To further investigate the relationship between newly defined medulloblastoma subtypes and neuronal differentiation, we integrated single-cell RNA-seq (scRNA-seq) data from cells at different stages of differentiation during normal development of the brain (Fig. 2a)¹². The Diffusion map, which represents the neuronal differentiation of the Rhombic lip with stemness characteristics, demonstrated two major axes of neuronal differentiation trajectories, granular neuron (GN) and unipolar brush cell differentiation (UBC). When we overlaid our medulloblastoma tumors onto the Diffusion map, we discovered that SHHa and SHH β tumors localized along the granular neuron differentiation axis, with SHH β exhibiting a higher degree of differentiation. In contrast, Group 3 tumors were located near the rhombic lip axis, reflecting their stem-like characteristics. Lastly, G4 α , G4 β , and G4 γ were aligned along the UBC differentiation axis, albeit transcriptome data did not distinguish among these subtypes.

Based on the Diffusion map, we calculated the proportion of neuronal cell types within each medulloblastoma subtype (Fig. 2b). The results showed that G3 tumors were exclusively composed of rhombic lip (RL)-type cells, whereas SHHa samples identified with granular cell progenitor (GCP)-type cells. Conversely, SHH β tumors included a mixture of GCP and differentiated granular neuron (GN)-type cells. The three Group4 subtypes showed a gradient pattern composed of both RL and UBC-type cells. To validate these findings, we used scRNA-seq data to calculate the proportions of cell lineage types in medulloblastoma tissues (Hendrikse et al., 2022) (Extended Data Fig. 8). The consistency in cell-type distributions across both datasets supports the robustness of our lineage tracing analyses.

Previous studies showed that medulloblastoma subgroups are characterized by unique expression patterns of transcription factors (TFs) known to regulate neuronal development^{13,14}. We visualized the transcriptional patterns of representative lineage-specific TFs on the Diffusion map (Fig. 2c and Extended Data Fig. 9). The TF profiles revealed distinct features for each medulloblastoma subtype: 1) OTX2-positive cells were aligned along the eCN/UBC axis and overlapped with Group 3 and Group 4 medulloblastoma tumors; 2) GLI2-positive cells were localized along the granular neuron axis and overlapped with SHH medulloblastoma tumors; 3) NEUROD1-positive cells were associated with non-WNT medulloblastoma tissues. We further examined the RNA and protein expression patterns of these lineage-specific TFs based on the new subtypes, identifying subtype-specific expression profiles (Fig. 2d). For instance, PAX3 characterized the WNT subgroup, GLI1 and GLI2 were predominantly active in the SHH subgroup, and OTX2 constituted the unique transcriptional profiles of Group 3 and 4.

As we found that NEUROD1-positive cells overlapped with non-WNT medulloblastoma tumors on the Diffusion map, we sought to explore the significance of this association. NEUROD1 is a well-known TF that regulates neuronal differentiation¹². Using publicly available ChIP-seq data, we defined NEUROD1 target genes, which were highly expressed in SHH β and G4 γ ($p = 0.0078$, Extended Data Fig. 10a). Next, we calculated correlation coefficients between NEUROD1 RNA and protein expression levels with its target genes (Extended Data Fig. 10b). Comparative analysis of the correlation coefficients between the neuronal subtype-related proteins and other target genes revealed that the abundance of the neuronal subtype-related proteins was significantly correlated with NEUROD1 expression compared to other target genes (Fig. 10b and c). Additionally, NEUROD1 RNA expression conferred favorable clinical outcomes in non-WNT medulloblastoma patients (Extended Data Fig. 10d). PFS analysis further showed that the expression of NEUROD1 mRNA and protein targets was significantly associated with improved prognosis in non-WNT medulloblastoma patients (Fig. 2e). These findings highlight the critical role of lineage-related TFs in medulloblastoma subtypes and suggest that medulloblastoma progression is linked to neuronal development process.

Proteogenomic analyses unveil targets and markers for SHH subtypes

The two newly defined subtypes, SHH α and SHH β , exhibited distinct clinical outcomes with SHH β patients demonstrating a significantly better prognosis compared to the SHH α group (Fig. 3a). Functional analysis showed that RNA and proteins implicated in cell cycle and DNA replication accumulated in the SHH α subtype, whereas neuronal proteins were expressed in the SHH β tumors. Moreover, the expression levels of proteins and corresponding transcripts within these functional categories were highly correlated (Fig. 3b and Extended Data Fig. 11). Specific cell cycle-related proteins, including CDK2, MCM2, and PARP1 were identified as representative SHH α markers (Fig. 3c and Extended Data Fig. 12). To determine if these proteins could serve as reliable biomarkers within clinical frameworks, we sought to validate their abundance in histology sections of medulloblastoma from tissue microarray (Fig. 3d and Extended Data Fig. 13). Consistent with our findings, a significant upregulation of CDK2, MCM2, and PARP1 was observed in the SHH α subtype (Fig. 3d).

Next, we sought to identify prognostic biomarkers associated with medulloblastoma progression that influence protein functions. To identify clinically relevant RNA biomarkers that correspond to protein functions, we selected genes with high RNA-protein correlations. Proteins related to neuronal function and cell cycle processes exhibited a high correlation, indicating transcriptional regulation of these proteins. Conversely, proteins associated with oxidative phosphorylation (OxPhos) and mitochondrial translation demonstrated poor RNA-protein correlations, suggesting post-transcriptional or translational regulation (Extended Data Fig. 11a). Interestingly, the integrative DNA methylation and RNA expression analysis revealed that neuronal function-related genes were highly associated with DNA methylation status, indicating potential epigenetic regulation of these genes (Extended Data Fig. 11b).

Using a Cox proportional hazard (coxph) model, we identified several core proteins with subtype-specific functions that were significantly associated with poor prognosis in medulloblastoma patients (Fig. 3e). We then leveraged the results of our correlation analysis to select RNA markers that affect protein functions (Extended Data Fig. 11c). Using selected RNA markers, we calculated a risk score using the coxph model and performed survival analysis. As a result, we discovered that 11 RNA markers representing cell cycle and metabolism categories were significantly associated with poor prognosis of SHH medulloblastoma patients (cell cycle: $p = 0.01$, $n = 21$; RNA metabolism: $p = 0.001$, $n = 21$) (Fig. 3e). To validate these findings, we analyzed the overall survival from a public dataset (GSE85217), which also showed that the same RNA markers were significantly associated with poor prognosis in SHH medulloblastoma patients (cell cycle: $p = 0.001$, $n = 198$; RNA metabolism: $p = 0.001$, $n = 198$) (Fig. 3e). Collectively, the prognostic subtype-specific markers identified through our integrated multi-omics analysis emerged as candidate biomarkers for clinical applications. They may also be useful to stratify medulloblastoma patients following identification of subtype-specific vulnerabilities.

Phosphoproteome analyses characterize SHHa-specific kinases

Our multi-omics approach integrated several layers of molecular profiles, including the phosphoproteome. We initially compared the phosphoproteome between SHHa and SHHb subtypes (Fig. 4a and b, Extended Data Fig. 14a). In the SHHa subgroup, both cell cycle and DNA replication-related phosphoproteins, such as CDK2 and EZH2 were markedly enriched, while the SHHb tumors exhibited enrichment of synapse-related phosphoproteins, including SYN1 and GABBR1. To identify subtype-specific phosphoproteins and upstream kinases, we performed a kinase activity-based phosphoproteomic analysis (Fig. 4c and d). Kinase–substrate enrichment analysis (KSEA) revealed distinct subtype-specific kinases and their substrate phosphoproteins (Fig. 4c). Using these representative kinase–substrate interactions, we constructed subnetwork modules highlighting the core subtype-specific kinases and their target substrates with their known functional roles (Fig. 4d). The resulting subnetworks were centered around key kinases, forming subtype-specific protein functional groups. Specifically, CDK1/2 and their substrates constituted a cell cycle-related module in the SHHa subtype, while CLK1 and its substrates composed an mRNA processing-related module. Next, we evaluated whether the identified functional modules were associated with medulloblastoma progression (Fig. 4e). We found that the phosphorylation of proteins modulated by subtype-specific kinases, such as CDK1/2 and CLK1 was significantly associated with PFS in each subgroup. These findings indicate that targeted inhibition of subtype-specific kinases may regulate the phosphorylation of key effector proteins, potentially altering disease progression and offering new therapeutic opportunities.

We further examined the multi-omic profiles of core proteins in SHHa medulloblastoma within known signaling pathways (Fig. 4f). Core proteins, including CDK1, CDK2, and PARP1, were predominantly involved in the cell cycle pathway. Interestingly, while most core proteins lacked somatic gene alterations, except for TP53, CCND2, and CDKN2A, they showed substantial RNA and protein expression changes, indicating transcriptional regulation of corresponding pathways. Given their continuous up-

regulation during disease progression and their enzymatic activities as kinase or DNA-damage response regulators, we selected CDK1, CDK2, and PARP1 as potential therapeutic targets for the SHHa subtype. Additionally, phosphorylation of CDK1 and PARP1 was consistently elevated during progression, indicating functional activation of these enzymes. We speculate that inhibition of these core kinases could disrupt the subtype-specific kinome and down-regulate related protein functions within the subnetwork. We also identified positive correlations between kinase protein expression and target substrate phosphorylation, further supporting the potential for targeted kinase inhibition to modulate subtype-specific signaling pathways (Fig. 4g and Extended Data Fig. 14b). These results highlight that subtype-specific kinase targeting may alter the phosphorylation state of effector proteins, which drive medulloblastoma progression and constitute promising therapeutic implications.

Synaptic signaling proteins are associated with good prognosis of medulloblastoma

Our multi-omics clustering and functional annotation revealed that proteins implicated in neuronal and synaptic functions accumulated at high levels in three medulloblastoma subtypes, SHH β , G4 β , G4 γ (Fig. 1a). Previous studies using proteomic data reported that SHH β tumors contained high levels of glutamatergic synaptic proteins and were more similar to tumors in Group 4 than those in the SHHa subtype⁹. Accordingly, our unsupervised clustering showed that the SHH β subtype clustered with Group 4 tumors, particularly the G4 γ subgroup, and neuronal proteins were enriched in both groups compared to G4 α and G4 β tumors (Fig. 5a). The neuronal proteins elevated in SHH β were primarily involved in synaptic signaling and were typically expressed in differentiated neural cells (Fig. 5a). Principal component analysis (PCA) further distinguished Group 4 patients into three subtypes, G4 α , G4 β , and G4 γ (Extended Data Fig. 15a). Interestingly, the two neuronal subtypes, SHH β and G4 γ , both enriched in synaptic signaling proteins, were significantly associated with favorable prognosis in non-WNT medulloblastoma (Extended Data Fig. 15b, $p = 0.015$ by log-rank test). Moreover, risk scores based on a coxph model using eleven synaptic transmission proteins showed a significant association with a better clinical outcome (Extended Data Fig. 15c, $p = 0.0006$ by log-rank test). Together, these findings suggest that tumor cells expressing synaptic transmission proteins belong to more differentiated and less aggressive tumors that may require more conservative treatment.

We also found that the expression of the proteasome complex was higher in the G4 β subtype, suggesting that differences among G4 α , G4 β , and G4 γ subtypes may arise not from transcriptional regulation but from post-transcriptional mechanisms, such as translation and proteasomal degradation of synaptic signaling proteins (Fig. 5a). Next, we sought to validate experimentally the above findings. Toward this aim, we first determined the proteogenomic subtypes of medulloblastoma cell lines using LC-MS/MS-based proteome analysis (Fig. 5b). By comparing the proteomic profiles of medulloblastoma cell lines with those of the seven medulloblastoma subtypes we have identified, we assigned each cell line to a specific subtype. Next, we investigated whether proteasome inhibition regulated synaptic signaling (Fig. 5c). Immunoblot assays showed that treatment with the proteasome inhibitor Bortezomib

regulated the expression of Synapsin-1 (Syn1), a protein marker of synaptic signaling and neuronal differentiation. Specifically, Syn1 was not expressed in the G4 α medulloblastoma cell line CHLA-01R-MED but was more abundant in the G4 β/γ cell line CHLA-01-MED. Upon treatment with Bortezomib, Syn1 further accumulated towards markedly higher levels, indicating that proteasome inhibition upregulated synaptic proteins. Finally, we assessed the therapeutic potential of proteasome inhibition in medulloblastoma. We measured the drug sensitivity of Bortezomib in the G4 β/γ medulloblastoma cell line (CHLA-01-MED) and showed that these cells were highly sensitive to nanomolar concentrations of Bortezomib (Fig. 5d). We measured the drug sensitivity of a proteasome inhibitor, Bortezomib, in the G4 β/γ medulloblastoma cell line (CHLA-01-MED). Our results demonstrated that the G4 β/γ medulloblastoma cell line was highly sensitive to nanomolar concentrations of Bortezomib. Using LC-MS/MS-based proteome analysis, we investigated expression patterns of synaptic signaling proteins in response to varying concentrations of Bortezomib in the G4 β/γ medulloblastoma cell line (Fig. 5e). Our analysis revealed two distinct expression profiles among synaptic signaling proteins. In one group, protein expression increased following proteasome inhibition by Bortezomib. Conversely, the second group displayed upregulation at lower Bortezomib concentrations but was downregulated at higher concentrations. These findings suggest that proteasome inhibition may restore synaptic signaling function, potentially leading to a less aggressive medulloblastoma subtype but it needs detailed treatment control for dose and concentration. However, precise control over Bortezomib dosage and concentration is essential to optimize treatment outcomes.

A receptor tyrosine kinase has therapeutic potential for Group 4 patients

To identify the underlying mechanisms sustaining Group 4 medulloblastoma, we focused on phosphoproteome data. First, we compared the phosphoproteomic landscape among G4 α , G4 β , and G4 γ subtypes and identified subtype-specific phosphorylation activities linked to distinct functional pathways (Fig. 5a). In the G4 α subtype, we found enrichments of proteins involved in EMT and receptor tyrosine kinase (RTK) signaling functions, including VIM, GSK3 β , and TGF β . The G4 β tumors exhibited activation of Myc pathway and cell cycle functions, including MDM1, SPP1, and ATRX. The G4 γ subtype was characterized by synaptic activity and neuronal differentiation functions, including SYN1, SYN2, SYT2, and MBP. To further delineate Group 4-specific phosphoproteins and their upstream kinases, we performed a kinase activity-based phosphoproteome analysis (KSEA) (Extended Data Fig. 16). By mapping representative kinase–substrate interactions, we constructed G4-specific subnetwork modules and annotated their functional roles (Extended Data Fig. 16). We found that G4 α subtype was driven by RTK-JAK-STAT signaling, with PDGFRB and JAK2 at its core, whereas G4 γ subtype was associated with neuronal functions, regulated by CAMKs, PRKCs, and MAPKs. To evaluate the therapeutic potential of targeting these pathways, we evaluated the pharmacological sensitivity of RTK inhibitors (Fig. 6d). In particular, we tested the effect of Foretinib, a Met kinase inhibitor, in the G4 α medulloblastoma cell line CHLA-01R-MED. These experiments showed that the G4 α medulloblastoma cell line was sensitive to

nanomolar concentrations of Foretinib, indicating that RTK inhibitors may be employed as promising therapeutic strategies for Group 4 (G4 α) medulloblastoma patients.

Medulloblastoma subtypes show different patterns of recurrence

Tumor recurrence remains therapeutically unresolved in medulloblastoma due to its complex evolutionary trajectory. To investigate its dynamic process, we conducted longitudinal molecular profiling of medulloblastoma by analyzing 10 cases with matched primary and recurrent tumor pairs (Extended Data Fig. 17). Initially, we compared the genome alterations between primary and recurrent tumors (Extended Data Fig. 17). Unsupervised hierarchical clustering and cell-cell correlation network analysis of longitudinal medulloblastoma revealed extensive inter-tumoral heterogeneity, defined by clustering of matched primary and recurrent tumors from the same patients (Extended Data Figs. 18 and 19). This finding indicates that the molecular difference between primary and recurrent tumor is marginal, and recurrence is not associated with a major subtype switch. Next, we categorized primary tumors into two groups: non-recurrent primary (NRP) tumors from patients without recurrence, and recurrent-potent primary (RPP) tumors from patients with recurrence. Using proteomic data, we calculated the average protein expression levels in NRPs, RPPs, and recurrent tumors and analyzed protein expression patterns that changed with recurrence (Extended Data Fig. 20). The results showed that most RPPs and their corresponding recurrent tumors exhibited consistent changes in protein expression, with upregulated proteins often being subtype-specific.

To investigate recurrence-associated changes across medulloblastoma subtypes, we analyzed recurrence-related protein functions (Fig. 6a and Extended Data Fig. 21). Each subtype of primary tumor was specifically enriched in specific protein functions, including cell cycle-related in the SHH α subtype, EMT and ECM reorganization functions in the G4 α subtype, and neuron-related functions were enriched in the G4 β and SHH β subtypes (Fig. 1). We specifically focused on the recurrence dynamics of SHH α and G4 β as they were the dominant recurrent subtypes. By comparing the protein functions between NRP and RPP, we identified recurrence-specific protein functions (Fig. 6a and Extended Data Fig. 21). In the SHH α subtype, the primary tumors were enriched in the cell cycle, mRNA processing and DNA repair functions, which became progressively more enriched as SHH α tumors progressed from NRP to recurrent tumors (Fig. 6a). In contrast, the G4 β subtype exhibited a shift from neuron-related functions in primary tumors to EMT and ECM activities in recurrent tumors, thus converting to G4 α -like phenotype. Next, we examined core protein expression dynamics in relation to primary and recurrence (Fig. 6b and c). We hypothesized two independent axes: the x-axis represents the expression of each core protein in the primary subtype, and the y-axis represents the expression changes at recurrence. Plotting core proteins related to cell cycle, EMT, and neuronal function based on expression values revealed that cell cycle functions, including the accumulation of CDK1 and MCM2, were enriched in the SHH α subtype, while the core proteins related to EMT functions, including the accumulation of POSTN and TGM2, were observed in the G4 β subtype (Fig. 6b and c). These results suggest that both EMT and neuronal functions are essential for recurrence in the Group 4 subgroup, and a deficiency in these functions may

need to be compensated during tumor relapse. Moreover, primary-recurrent pairs were clustered together, indicating that the neuronal cell types remained stable based on transcriptome data.

Finally, we sought to experimentally validate our findings in two Group 4 medulloblastoma cell lines: G4 β -like (or G4 γ -like), CHLA-01-MED and G4 α -like, CHLA-01R-MED. The CHLA-01-MED cell line was characterized by neuronal and synaptic signaling functions, whereas the CHLA-01R-MED cell line was characterized by EMT-associated characteristics. Notably, these two cell lines are primary-recurrence pairs derived from the same patient. Constructing a cell-to-cell interaction network with proteome data allowed us to further refine the subtyping of medulloblastoma cell lines (Extended Data Fig. 23). To uncover the therapeutic effects of MET kinase inhibition shown in Fig. 6d, we performed LC-MS/MS-based proteome analysis after treating the medulloblastoma cell lines with Foretinib (Fig. 6d, E and Extended Data Fig. 24). Strikingly, our results highlighted a dual function of Met inhibition on Group 4 medulloblastoma. In the aggressive G4 α -like medulloblastoma cells, EMT function were suppressed. Moreover, synaptic signaling function, which was initially low in G4 α -like cells, was restored to the levels observed in G4 β - or G4 γ -like medulloblastoma cells.. This suggests that Met inhibition not only reduced EMT activity but also promotes differentiation towards a neuronal subtype. Consequently, we propose that targeting proteome and kinome to modulate neuronal differentiation represents a promising therapeutic strategy for treating medulloblastoma.

Discussion

Many omics studies have classified medulloblastoma into four subgroups, WNT, SHH, Group 3 and Group 4, which currently represent the global standard of medulloblastoma classification⁴. By integrating proteomics data with other existing omics data, several reports described new subtypes, each of which is sustained by subtype-related pathways and activities proposed as potential therapeutic targets^{9,15}. In spite of many attempts to integrate proteomics data, the existing global standard has not been improved. This is due to relatively limited numbers of samples, the scale of the multiomics analysis platform, and inconsistency in analysis methods. We generated datasets from five omics platforms, including proteomics data from ~ 120 tumor samples of medulloblastoma from ~ 90 patients. In our study, the numbers of patients, samples and data types are larger than those of previous proteomics-integrating medulloblastoma omics study. Moreover, our sample collection contains not only primary tumors but also matched recurrent tumors. We compared our results to those of other medulloblastoma studies, validated their results and identified new perspectives and types of data with more statistical power than other proteogenomic studies.

Our findings enhance the reliability of existing results. First, our unsupervised clustering of methylomic and transcriptomic data well matched the conventional four subgroups, WNT, SHH, and Groups 3 and 4 (Extended Data Figs. 4 and 5). By integrating proteomic data, we subdivided the SHH subgroup into SHH α and SHH β subtypes. Archer *et al.* also subdivided the SHH subgroup into SHH α and SHH β subtypes, which showed similar functional enrichment in the proteome, such as RNA processing, MYC targets and DNA repair in SHH α , and synaptic signaling and axon guidance in SHH β ⁹. They also found

that SHH β was closer to Group 4 than SHH α in proteome clustering. Similarly, according to our unsupervised clustering of proteomic data, SHH β was most closely related to the G4 β subtype (Extended Data Fig. 2). The similarity between our data and those reported in previous studies provides an accurate independent validation to our new findings.

We were able to extend the scope of our investigation to recurrence in medulloblastoma as we generated omics data from matched recurrent tumor tissues. Although the prognosis of relapsed patients is substantially poorer than for patients with primary tumors, most proteogenomic studies of medulloblastoma have been restricted to primary tumors. By analyzing the functions that were enriched during recurrence, we found that the trends in protein functions were different among subtypes, especially between SHH and Group 4 (Fig. 6). In the SHH α subtype, the protein functions of primary tumors were enriched in cell cycle-related functions, including G2/M checkpoint, DNA replication and repair, and these functions are further enriched during recurrence. We note that our study analyzed markedly more samples classified as Group 4 medulloblastoma than were available in the study of Archer *et al.*. Therefore, we were able to examine Group 4 more closely and subdivide it into the G4 α , G4 β and G4 γ subtypes. The protein expression profiles of G4 α showed a significant enrichment for EMT-related functions, while G4 β and G4 γ proteomes were enriched in neuron-related functions. At recurrence, the functional proteomic profile of G4 β switched to the activation of EMT-related functions (Fig. 6). From these findings, we suggest medulloblastoma recurrence is marked by a convergence towards mesenchymal transition.

Additionally, we compared the average gene expression differences between primary and recurrent tumor samples in medulloblastoma and glioblastoma (Extended Data Fig. 25). The left half of the plot represents medulloblastoma, while the right half represents glioblastoma (GBM). Each dot represents the average difference in gene expression levels (absolute values) between primary and recurrent tumor samples for individual genes. The radial distance from the center indicates the magnitude of the difference, with greater distances representing larger changes in average expression. The colors differentiate between the two tumor types, with medulloblastoma shown in green and GBM in red. The circular layout visually compares the expression variability and the magnitude of gene expression alterations between the two tumor types.

Performing lineage tracing analysis with public scRNA-seq data, we have identified that medulloblastoma subtypes display two distinct lineage-specific neuronal differentiation patterns: granular neuron differentiation associated with the SHH subtype and UBC differentiation linked to Group 4. Along the granular neuron differentiation axis, the SHH β subtype with more favorable prognostic outcomes, exhibits a higher degree of neuronal differentiation compared to SHH α . The transcriptional regulatory mechanisms involve well-established lineage-related TFs, notably NEUROD1. In contrast, along the UBC differentiation axis, transcriptomic data alone cannot distinctly separate the three subtypes of Group 4. However, proteomic and kinomic features successfully differentiate G4 γ with better prognostic outcomes, from G4 α and G4 β . Similar to SHH β , G4 γ shows a higher degree of neuronal differentiation compared to G4 α and G4 β , but the regulatory mechanisms involve proteomic and kinomic

features, such as proteasome activity and RTK signaling. This suggests that multiomic regulatory compensation, rather than transcriptome alone, is pivotal for neuronal differentiation and prognosis, highlighting the importance of proteome and kinome in these processes.

Proteins are more functionally meaningful molecules than DNA or RNA because they directly serve as enzymes, receptors, structural elements like the cytoskeleton, and more. If alterations in DNA or RNA are not linked to protein functions, they may not drive phenotypic changes. In comparing the G4 β and G4 γ subtypes, we did not identify significant transcriptomic differences. However, several synaptic signaling proteins, which are highly expressed in SHH β , characterize the G4 γ subtype in Group 4. These synaptic signaling proteins promote neuronal differentiation marker proteins, including Syn1. Synaptic signaling proteins are downregulated except in the G4 γ subtype, and significantly associated with good prognosis for medulloblastoma patients. Similarly, proteasome activity is a regulatory mechanism that cannot be detected through genomic or transcriptomic analyses alone. This highlights the importance of proteogenomic analyses in diagnosis, as they can reveal critical regulatory processes that are overlooked by other methods.

Another important aspect of mass spectrometry-based proteome analysis is its ability to investigate post-translational modifications (PTM). The G4 α subtype, known for its poor prognosis, can be characterized by global proteome and phosphoproteome analyses, revealing EMT functions and RTK signaling pathways. By employing activity-based phosphoproteomic analysis and kinase–substrate network, we identified subtype-specific kinases and corresponding inhibitors targeting a central kinase within a kinome module. We demonstrated that these selected kinase-specific inhibitors can effectively inhibit medulloblastoma cell growth in a subtype-specific manner. Furthermore, our proteome analyses revealed a dual therapeutic mechanism of Met kinase inhibition. This mechanism not only inhibits EMT, a critical process in tumor progression, but also enhances synaptic signaling recovery, thereby promoting a less aggressive neuronal subtype of medulloblastoma. This highlights the utility of activity-based phosphoproteomic analysis in therapeutic target selection.

Proteome-based prognostic markers and potential therapeutic targets can be strategically paired. For instance, when RNA markers related to cell cycle functions predict a poor prognosis in SHH medulloblastoma patients, Cdk1 and Cdk2 inhibitors can be beneficial. Similarly, when EMT function-related protein markers indicate a poor prognosis in Group 4 medulloblastoma patients, Met inhibitors are advantageous. Given that many small-molecule drugs and libraries targeting kinases have been developed—and that most kinases can be targeted with known drugs—our method for identifying prognostic markers and specific kinase inhibitor targets can be highly effective for repositioning known kinase inhibitors. Through multiomics analysis, we identified several subtype-specific prognostic marker–therapeutic target combinations. We suggest that multiomics approaches, including proteomic and phosphoproteomic analyses, will accelerate diagnostics and therapeutics by enhancing the efficiency of drug and marker development.

Methods

Tissue cryopulverization

The frozen tissue samples were weighed and washed with cold phosphate-buffered saline (PBS) to remove contaminating blood, placed into tissueTUBEs (Covaris, Woburn, MA), snap-frozen in liquid nitrogen and pulverized using a cryoPREP Tissue Disruption system (CP02, Covaris). The pulverized tissue powder was aliquoted (10–20 mg) for DNA, RNA and protein extraction.

DNA and RNA extraction and quality analysis

In the case of frozen tissues, genomic DNA was extracted from pulverized tissues using a MagNA Pure 24 Total NA Isolation kit with a MagNAPure 24 automated instrument (Roche, Switzerland) or Allprep DNA/RNA Mini Kit (Qiagen, Germany). In the case of blood samples, genomic DNA was extracted from buffy coat using Maxwell® 16 Blood DNA Purification Kit with Maxwell® 16 automated instrument (Promega, USA) or using Maxwell® RSC Buffy Coat DNA Kit with a Maxwell® RSC 48 automated instrument (Promega, USA) or using a QIAamp DNA Blood Mini kit (Qiagen, Germany). Purified gDNA was analyzed for concentration and purity with a Nanodrop 8000 (Thermo Fisher, USA). The integrity of the gDNA was analyzed with 1% gel electrophoresis. DNA quantitation was analyzed using a Qubit dsDNA BR Assay Kit (Thermo Fisher, USA).

In the case of frozen tissues, total RNA was extracted from pulverized tissues using an Allprep DNA/RNA Mini Kit with a QIAcube automated instrument (Qiagen, Germany) or an RNeasy Mini Kit with a QIAcube automated instrument (Qiagen, Germany). Purified RNA was analyzed for concentration and purity with a Nanodrop 8000 (Thermo Fisher, USA). For RNA quality analysis, RNA was analyzed using an RNA Nano 6000 Kit with a Bioanalyzer 2100 (Agilent, USA).

Whole-exome sequencing

The SureSelect Target Enrichment workflow is a solution-based system that utilizes ultralong 120-mer biotinylated cRNA baits to capture regions of interest, enriching them from an NGS genomic fragment library. To generate standard exome capture libraries, we used the Agilent SureSelectXT Low Input Target Enrichment protocol for the Illumina paired-end sequencing library with 1 µg of input gDNA. The DNA quantity and quality were measured by PicoGreen and agarose gel electrophoresis, respectively. We used 1 µg of each genomic DNA diluted with EB Buffer and sheared to a target peak size of 150–200 bp with the Covaris LE220 focused-ultrasonicator (Covaris, Woburn, MA) according to the manufacturer's recommendations. Fragmentation is followed by end repair and the addition of an 'A' tail. Agilent adapters were then ligated to the fragments. After assessing the ligation efficiency, the adapter-ligated product was PCR amplified. The final purified product was quantified with the TapeStation DNA ScreenTape D1000 platform (Agilent). For exome capture, 250 ng of DNA library was mixed with hybridization buffers, blocking mixes, RNase block and 5 µl of SureSelect all exon capture library, according to the standard Agilent SureSelect Target Enrichment protocol. The captured DNA was

washed and amplified. Then, the final purified product was quantified by qPCR according to the qPCR Quantification Protocol Guide (KAPA Library Quantification kits for Illumina Sequencing platforms) and tested for quality with the TapeStation DNA ScreenTape D1000 platform (Agilent). Illumina utilizes a unique amplification reaction that occurs on the surface of the flow cell. A flow cell containing millions of unique clusters is loaded into the Illumina platform for automated cycles of extension and imaging. Sequencing-by-synthesis utilizes four proprietary nucleotides possessing reversible fluorophore and termination properties. Each sequencing cycle occurs in the presence of all four nucleotides, leading to higher accuracy than methods where only one nucleotide is present in the reaction mix at a time. This cycle is repeated one base at a time, generating a series of images, each representing the extension of a single base at a specific cluster. The Illumina platform generates raw images and performs base calling through its integrated primary analysis software, RTA (Real Time Analysis). The base calling files, which are in binary format, were converted into FASTQ by the Illumina package bcl2fastq v2.20.0. The demultiplexing option (--barcode-mismatches) was set to 0.

Whole-genome sequencing

The samples were prepared according to the Illumina TruSeq Nano DNA library preparation guide. The libraries were sequenced using the Illumina NovaSeq6000 platform. Each sequenced sample was prepared according to the Illumina TruSeq Nano DNA sample preparation guide to obtain a final library with an average insert size of 300–400 bp. One hundred nanograms of genomic DNA was fragmented by the Covaris system, which generates dsDNA fragments with 3' or 5' overhangs. The dsDNA fragments with 3' or 5' overhangs were converted into blunt ends using an end repair mix. The 3' to 5' exonuclease removed the 3' overhangs, and the polymerase filled the 5' overhangs. Following end repair, the appropriate library size was selected using different ratios of the sample purification beads. A single 'A' nucleotide was added to the 3' ends of the blunted fragments to prevent them from ligating to one another during the adapter ligation reaction. A corresponding single 'T' nucleotide on the 3' end of the adapter provides a complementary overhang for the adapter to be ligated to the fragment. Multiple indexing adapters be ligated to the ends of the DNA fragments to prepare them for hybridization onto a flow cell. PCR was used to amplify the enriched DNA library for sequencing. PCR was performed with a PCR primer solution that anneals to the ends of each adapter. Quality control analysis of the sample library and quantification of the DNA library templates was performed by MacroGen. For cluster generation, the library was loaded into a flow cell, where the fragments were captured on a lawn of surface-bound oligos complementary to the library adapters. Each fragment was then amplified into distinct, clonal clusters through bridge amplification. After cluster generation, the templates were submitted for sequencing. Illumina SBS technology utilizes a proprietary reversible terminator-based method that detects single bases as they are incorporated into DNA template strands. As all 4 reversible, terminator-bound dNTPs are present during each sequencing cycle, natural competition minimizes incorporation bias and greatly reduces the raw error rates relative to those of other technologies. The result is highly accurate base-by-base sequencing that virtually eliminates sequence-context-specific errors, even within repetitive sequence regions and homopolymers. The Illumina Platform generates raw

images and performs base calling through an integrated primary analysis software called RTA (Real Time Analysis). The BCL/cBCL (base calls) binary is converted into FASTQ using the Illumina package bcl2fastq2-v2.20.0. The demultiplexing option (`--barcode-mismatches`) was set to a perfect match (value: 0).

Whole-transcriptome sequencing

The total RNA concentration was calculated by Quant-IT RiboGreen (Invitrogen, #R11490). To assess the integrity of the total RNA, samples were run on the TapeStation RNA ScreenTape platform (Agilent, #5067–5576). Only high-quality RNA preparations with RIN values greater than 7.0 were used for RNA library construction.

A library was independently prepared with 1 µg of total RNA for each sample by an Illumina TruSeq Stranded mRNA Sample Prep Kit (Illumina, Inc., San Diego, CA, USA, #RS-122-2101). The first step in the workflow involves purifying the poly-A-containing mRNA molecules using poly-T-attached magnetic beads. Following purification, the mRNA was fragmented into small pieces using divalent cations under elevated temperature. The cleaved RNA fragments were copied into first-strand cDNA using SuperScript II reverse transcriptase (Invitrogen, #18064014) and random primers. This was followed by second-strand cDNA synthesis using DNA Polymerase I, RNase H and dUTP. These cDNA fragments were then subjected to an end-repair process, the addition of a single 'A' base, and adapter ligation. The products were then purified and enriched with PCR to create the final cDNA library.

The libraries were quantified using KAPA Library Quantification kits for Illumina Sequencing platforms according to the qPCR Quantification Protocol Guide (KAPA BIOSYSTEMS, #KK4854) and qualified using the TapeStation D1000 ScreenTape platform (Agilent Technologies, # 5067–5582). Indexed libraries were then paired-end (2×100 bp) sequenced with an Illumina NovaSeq (Illumina, Inc., San Diego, CA, USA) at MacroGen Incorporated.

Methylome

DNA samples were checked for quality using a NanoDrop® ND-2000 UV–Vis spectrophotometer. Then, the samples were electrophoresed on agarose gels, and samples with intact genomic DNA showing no smearing on agarose gel electrophoresis were selected for the experiment. Intact genomic DNA was diluted to 50 ng/µl based on Quant-iT Picogreen (Invitrogen) quantitation. All prepared samples were bisulfite-converted according to the Zymo EZ DNA methylation kit protocols.

A total of 600 ng of input gDNA was required for bisulfite conversion. The conversion reagent was added, and the reaction mixture was incubated in a thermocycler for denaturation. CT-converted DNA was washed and desulfonated with desulfonation buffer. After desulfonation, the DNA was washed again and eluted in 12 µl elution buffer.

The whole-genome amplification process requires 250 ng of input bisulfite-converted DNA, MA1, which creates a sufficient quantity of DNA (1000X amplification) to be used on a single BeadChip in the Infinium methylation assay (Illumina RPM and MSM). After amplification, the product was fragmented using a proprietary reagent (FMS), precipitated with 2-propanol (plus precipitating reagent; PM1), and resuspended in formamide-containing hybridization buffer (RA1). The DNA samples were denatured at 95°C for 20 min and then placed in a humidified container for a minimum of 16 h at 48°C, allowing the CpG loci to hybridize to the 50-mer capture probes.

Following hybridization, the BeadChip/Te-Flow chamber assembly was placed on the temperature-controlled Tecan Flowthrough Chamber Rack, and all subsequent washing, extension, and staining steps were performed by adding the reagents to the Te-Flow chamber.

For the allele-specific single-base extension assay, primers were extended with a polymerase and labeled nucleotide mix (TEM) and stained with repeated application of STM (staining reagent) and ATM (anti-staining reagent). After staining was complete, the slides were washed with low-salt wash buffer (PB1), immediately coated with XC4, and then imaged on the iScan System.

Protein extraction and peptide digestion

Tissue powder samples were solubilized in SDS solubilization buffer (5% SDS, 50 mM TEAB pH 8.5) using an S220 Focused-ultrasonicator (Covaris). Proteins were digested using S-Trap™ spin columns (Protifi, Huntington, NY) according to the manufacturer's instructions. The samples were reduced by DTT and alkylated by iodoacetamide (IAA). After quenching the alkylation reaction, additional SDS and phosphoric acid were added so that the final concentration was 5% SDS and 1.2% phosphoric acid. The acidified samples were mixed with 90% methanol in 100 mM TEAB, loaded into S-Trap micro columns, and incubated with mass spectrometry-grade trypsin/LysC (Promega) for 3 h at 47°C. The eluted peptides were evaporated using a vacuum concentrator and cleaned up using C18 spin columns (Thermo Fisher Scientific, Rockford, IL).

TMT 11-plex labeling

Desalted peptide samples were reconstituted in 100 mM TEAB and labeled using TMT 11-plex reagents (Thermo Fisher Scientific). Each prepared TMT reagent was transferred to the peptide sample, and the mixture was incubated for 1 h, quenched by the addition of 8 µL of 5% hydroxylamine and then incubated for a further 15 min at room temperature. Differently labeled 11-plex peptides were pooled and dried using a vacuum concentrator.

Peptide fractionation by mid-pH reversed-phase liquid chromatography

The pooled 11-plex TMT-labeled sample was separated using an Agilent 1260 Infinity HPLC system (Agilent, Palo Alto, CA). An Xbridge C18 analytical column (4.6 mm × 250 mm, 130 Å, 5 µm) and a guard column (4.6 mm × 20 mm, 130 Å, 5 µm) were used for peptide separation. Solvents A and B were 10 mM triethylammonium bicarbonate (TEAB) in water (pH 7.5) and 10 mM TEAB in 90% acetonitrile (ACN, pH 7.5), respectively. Peptide fractionation was performed using a 120 min gradient at a flow rate of 500 µL/min as follows: 0% solvent B for 15 min, 0 to 5% solvent B over 10 min, from 5 to 35% solvent B over 60 min, from 35 to 70% solvent B over 15 min, 70% solvent B for 10 min, and from 70 to 0% solvent B over 10 min. A total of 96 fractions were collected, with one collected every minute from 15 to 110 min, and pooled into 24 noncontinuous peptide fractions (i.e., #1–#25–#49–#73, #2–#26–#50–#74, ..., #24–#48–#72–#96) and dried using a concentrator.

Phosphopeptide enrichment using immobilized metal affinity chromatography (IMAC)

Ni-NTA agarose beads (Qiagen, Valencia, CA) were washed with deionized water and treated with 100 mM EDTA, pH 8.0, for 30 min with end-over-end rotation. The EDTA solution was removed, and the beads were washed with deionized water and treated with 10 mM aqueous FeCl₃ metal ion solution for 30 min with end-over-end rotation. After the removal of excess metal ions, the beads were washed with deionized water, resuspended in 1:1:1 acetonitrile/methanol/0.01% acetic acid solution and aliquoted into microcentrifuge tubes. Fractionated peptide samples were resuspended in resuspension buffer (80% acetonitrile, 0.1% TFA). After washing the aliquoted Fe³⁺-NTA beads with resuspension buffer, the resuspended peptide sample was added and incubated with end-over-end rotation for 30 min. The supernatant was collected and dried for future analysis, the beads were washed with resuspension buffer, and the remaining solution was discarded. The enriched phosphopeptide was eluted with 1:1:1 acetonitrile/2.5% ammonia–water/2 mM phosphate buffer, acidified with 10% TFA solution and dried using a vacuum concentrator.

LC–MS/MS analysis

TMT-labeled peptides prepared for global proteome and phosphoproteome analysis were resuspended with 0.1% formic acid in water, separated using an Ultimate 3000 RSLCnano system (Thermo Fisher Scientific, San Jose, CA, USA) and analyzed using a Q Exactive HF-X hybrid quadrupole-Orbitrap mass spectrometer or a Q-Exactive plus hybrid quadrupole Orbitrap mass spectrometer (Thermo Fisher Scientific, San Jose, CA, USA).

For the Q Exactive HF-X hybrid quadrupole-Orbitrap mass spectrometer, solvents A and B were 0.1% FA in water and 0.1% FA in acetonitrile, respectively. The peptides were loaded onto the trap column (Accluum PepMap™ 100, 75 µm x 2 cm), separated by the analytical column (EASY-Spray column, PepMap™ RSLC C18, 75 µm x 50 cm, Thermo Fisher Scientific) with a gradient of 5 to 24% solvent B for

150 min and 24 to 36% solvent B for 30 min (for the global proteome) or 5 to 24% solvent B for 170 min and 24 to 36% solvent B for 10 min (for the phosphoproteome) at a flow rate 0.3 μ L/min. The Q Exactive HF-X Orbitrap mass analyzer was operated in a top 10 data-dependent method. Full MS scans were acquired over the range m/z 350–2000 with a mass resolution of 120,000 (at m/z 200). The AGC target value was $3.00E + 06$. The ten most intense peaks with a charge state ≥ 2 were fragmented in the higher-energy collisional dissociation (HCD) collision cell with a normalized collision energy of 32, and tandem mass spectra were acquired in the Orbitrap mass analyzer with a mass resolution of 45,000 at m/z 200.

For a Q-exactive plus hybrid quadrupole Orbitrap mass spectrometer, peptides were loaded from an RS autosampler and separated with a linear gradient of ACN/water containing 0.1% formic acid with a flow rate of 300 nL/min. Chromatographic separation of peptides was achieved using the Ultimate 3000 RSLC nano system equipped with the same column setting with HF-X. The LC eluent was electrosprayed directly from the analytical column, and a voltage of 2.0 kV was applied via the liquid junction of the nanospray source. Peptide mixtures were separated with a stepwise gradient from 7–60% ACN over 105 min (7–25% min for 85 min & 25% to 60 min for 20 min). The analysis method consisted of a full MS scan with a range of 350–2000 m/z and data-dependent MS/MS (MS2) on the ten most intense ions from the full MS scan. The mass spectrometer was programmed to acquire data in data-dependent mode. The mass spectrometer was calibrated with the proposed calibration solution according to the manufacturer's instructions.

Genomic data analysis

Sequencing reads were aligned to the GRCh37/hg19 human reference genome using the Burrows–Wheeler Aligner (BWA) and further processed by GATK to remove low mapping quality reads and realign sequences around indels. To confirm that tumor and blood samples were derived from the same patient, we performed fingerprint analysis using NGSCheckMate, a model-based method that evaluates the correlation of the variant allele fractions at known SNP sites. Somatic SNVs and indels were identified by integrating the results from 3 variant calling algorithms: TNhaplotyper, TNscope and TNSnv. For tumor samples without matched blood DNA, the somatic status of called mutations was assessed using a virtual normal panel from a set of 433 public samples from healthy, unrelated individuals sequenced to high depth in the context of the 1000 Genomes Project. Putative false positive calls were removed by applying a previously described in-house filtering pipeline. The mutation profiles of the longitudinal medulloblastoma pairs (10 primary and 10 recurrent matched samples) were built by combining the genetic variants identified in the two biopsies from each single patient, each of which was annotated as shared (occurring in both primary and recurrent samples) and primary or recurrent private (occurring exclusively in the primary or in the recurrent sample, respectively). To validate the sharing profile of mutations, the nucleotide at each mutant position was re-called from the raw sequences within both primary and recurrent samples from a single patient. Using this iterative approach, false negative calls were retrieved by identifying mutant reads at genetic positions that had been mis-called as wild type.

Somatic variants were annotated using AnnoVar, which aggregates information from genomic and protein resources in cancer and noncancer variant databases. Variants reported in the noncancer databases with a minor allele frequency ≥ 0.05 were classified as germline polymorphisms and excluded.

The functional effects of missense SNVs and in-frame indels were determined by an ensemble of multiple algorithms. Variants predicted as damaging by two or more algorithms were classified as pathogenic mutations.

Somatic copy number was estimated from WES and WGS by CNVkit. GISTIC2 analysis was then applied to integrate the results from individual patients and identify genomic regions that were recurrently amplified or deleted in the medulloblastoma cohort.

Transcriptomic data analysis

Raw fastq files were processed for adapter trimming using the Cutadapt program (ver. 2.9, <https://cutadapt.readthedocs.io/en/stable/index.html>). The processed files were aligned to the human reference genome (GRCh38) using the HISAT2 program (ver. 2.1.0, <https://github.com/DaehwanKimLab/hisat2>). The aligned sam files were converted to bam files and sorted by coordinate using the Samtools program (ver. 1.10, <http://www.htslib.org>). Duplicate reads were removed using Picard (ver. 2.22.1, <https://github.com/broadinstitute/picard>). The read counts for gene expression were computed using the HTseq program (ver. 0.11.3, <https://htseq.readthedocs.io/en/master/>). The gene expression values were calculated based on the fragments per kilobase of exon per million (FPKM) value. Genes that had greater than 30% missing values were discarded. The expression levels of the filtered genes were globally normalized with the Quantile normalization method using the R (ver. 4.0) limma package or MATLAB Bioinformatics Toolbox (ver. R2021a).

Methylomic data analysis

For scanning the Illumina 850k EPIC microarray, we used the iScan System, which is a two-color (532 nm/658 nm) confocal fluorescent scanner with 0.54 μm pixel resolution. The scanner excites the fluorophores generated during signal amplification/staining of the allele-specific (one-color) extension products on the BeadChips. The image intensities were extracted using Illumina's GenomeStudio software.

Proteomic data analysis

Raw files of tandem mass spectra were converted into mzML files using the msConvert program (ver. 3.0, <https://bio.tools/msconvert>). The mzML spectral data were mapped to the human UniProt database (UP000005640.fas, <https://www.uniprot.org/>) and quantified using the FragPipe pipeline (ver. 14.0),

including MSFragger, Philosopher and TMT-integrator program (ver. 3.1.1, <https://msfragger.nesvilab.org>). All identified proteins had an FDR of $\leq 1\%$, which was calculated at the peptide level. For the global proteome, the search parameters allowed for tryptic specificity of up to two missed cleavages, with methylthio-modifications of cysteine as a fixed modification and oxidation of methionine as a dynamic modification. The mass search parameters for -1 , 0 , $+1$, $+2$, and $+3$ ions included mass error tolerances of 20 ppm for precursor ions. For the phosphoproteome, the search parameters allowed for tryptic specificity of up to two missed cleavages, with modified serine, tyrosine, and threonine as variable modifications. The mass search parameters for 0 , $+1$, $+2$, and $+3$ ions included mass error tolerances of 20 ppm for precursor ions. The protein expression was quantified with the isobaric TMT 11 option and normalized to the relative expression ratio against the global reference pool of each TMT 11 set. Protein sequences with greater than 30% missing values were discarded. The expression ratios of filtered proteins were globally normalized with the Quantile normalization method using the R (ver. 4.0) limma package or MATLAB Bioinformatics Toolbox (ver. R2021a).

Unsupervised clustering

For single-platform data clustering, we used two unsupervised clustering methods, hierarchical clustering and nonnegative matrix factorization (NMF) clustering. Hierarchical clustering was performed using the MEV program (ver. 4.9, <https://sourceforge.net/projects/mev-tm4/>). NMF clustering was performed using the MATLAB Bioinformatics Toolbox (ver. R2021a). The mean absolute deviation (MAD) of each gene was calculated, and the genes with the top 10%, 20%, and 30% MADs were selected as the core genes. By changing the MAD cutoff and the number of clusters (k -value = $2 \sim 8$), the cophenetic correlation coefficient and silhouette coefficient were calculated, and the clustering was optimized based on the determined coefficients.

For multiomic clustering, we constructed a binary matrix of the single-platform cluster to which each sample belonged (Extended Data Fig. 3). With the binary matrix, hierarchical clustering was performed using the MEV program. The alluvial plot was generated with the ggplot2 (ver. 3.3.5) package in R.

Selection of differential expression and functional annotation

For the transcriptomic and global proteomic data, differentially expressed genes and proteins were selected by Gene Set Enrichment Analysis (GSEA) (ver. 4.1.0, <https://www.gsea-msigdb.org/gsea/index.jsp>). The functions of core genes and proteins were annotated using GSEA and DAVID (<https://david.ncifcrf.gov/>).

For the phosphoproteomic data, the selection of differentially expressed phosphoproteins and quantitation of kinase activity were performed using the Kinase–Substrate Enrichment Analysis (KSEA)

program (<https://casecpb.shinyapps.io/ksea/>). The kinase–substate network was generated using the Cytoscape program (ver. 3.8.0, <https://cytoscape.org>).

The heatmaps were drawn using Excel (Microsoft) and MEV. The density plots were drawn using the ggplot2 package in the R program.

Pathway analysis

Core pathways, genes, and proteins were selected from the Reactome (<https://reactome.org/>) and KEGG (<https://www.genome.jp/kegg/>) pathways in MSigDB (<https://www.gsea-msigdb.org/gsea/msigdb>). The pathways and multiome heatmaps were manually drawn using Excel and PowerPoint (Microsoft).

Tissue microarray and immunohistochemistry

We reviewed hematoxylin–eosin slides and selected three representative areas per slide, often with high mitotic counts, while avoiding necrotic and hemorrhagic regions. Tissue cores (diameter, 2 mm) obtained from these areas were then inserted into recipient TMA blocks (SuperBioChips, Seoul, Republic of Korea).

IHC staining was performed on 2-µm-thick formalin-fixed and paraffin-embedded (FFPE) tissue microarray slides using an automated immunostaining system (BenchMark GX system; Ventana-Roche, Mannheim, Germany). The primary antibodies used in this study are summarized in Supplementary Table XX. For the positive control, internal positive control tissue or cells were used, and for the negative control, primary antibodies were omitted.

Proliferation assay

Cells were plated at a density of 3×10^3 cells/well in 96-well plates containing DMEM (10% FBS, 1x P/S) for in vitro proliferation assays. The luminescence of viable cells was detected 2 days after plating using a CellTiter-Glo Luminescent Cell Viability Assay Kit according to the manufacturer's protocol (Promega). The luminescence signal was detected by a SpectraMax L Microplate Reader (Molecular Device) according to the manufacturer's protocol.

Declarations

A statement of ethics approval:

This study is approved by the IRB of Seoul National University Hospital (H-1805-061-945).

Data availability

Raw omics data have been deposited in public repositories. Proteomic data have been deposited in the Proteomic Data Commons (PDC, <https://proteomic.datacommons.cancer.gov/pdc/edu/>; ID PDC000522, PDC000523, PDC000524, PDC000525). DNA methylation idat files have been deposited to the Gene Expression Omnibus (GEO; ID GSE209668). Raw sequencing (fastq) files have been deposited into the SRA database (<https://www.ncbi.nlm.nih.gov/sra>) under project ID PRJNA862984, PRJNA863327, PRJNA864070 and PRJNA865394.

Declaration of interests

The authors declare no competing interests.

Author contributions

J.B.P., S.-K.K., S.-H.P., J.K.S, Y.W.K, J.T.K, H.Y, H.-S.G, M.D.T and A.I. conceived this study. S.-M.P., K.-H.K., J.H.Y., S.A.C. and J.B.P. designed the experiments. K.-H.K., J.H.Y., C.I.K, S.A.C., S.S.K., Y.M.S., Y.S.J., H.J.K, S.J.H, S.H.P, H.J.S and D.H.N. performed the experiments. S.-M.P., S.M.P, H.D.K, S.R.S, H.R.K, F.D.A., S.P, E.J.K., S.-I.K., K.-H.L., A-K.P., Y.K., J.K.S, Y.W.K, D.H.H., D.Y.H, S.H.H and J.H.P. analyzed the data with input from J.B.P., S.-K.K., S.-H.P. and A.I., S.-M.P., K.-H.K., F.D.A., J.H.Y., A.I., S.-H.P., S.-K.K. and J.B.P. wrote the manuscript.

Acknowledgments

This work was done under the auspices of a Memorandum of Understanding between National Cancer Center of Korea (NCC) and the U.S. National Cancer Institute's International Cancer Proteogenome Consortium (ICPC). ICPC encourages international cooperation among institutions and nations in proteogenomic cancer research in which proteogenomic datasets are made available to the public. We thank Dr. Henry Rodriguez and Dr. Ana I. Robles from the U.S. National Cancer Institute's Clinical Proteomic Tumor Analysis Consortium (CPTAC) for helpful discussions.

This work was supported by the National Cancer Center Grant (NCC-1810861), National Research Foundation of Korea (NRF) grant (2021R1A2C301331511 and 2021M3F7A108323011), SNUH Kun-hee Lee Child Cancer & Rare Disease Project, Republic of Korea (22A-017-0100) and KBRI basic research program through Korea Brain Research Institute (22-BR-02-03) funded by the Korean government (MSIT).

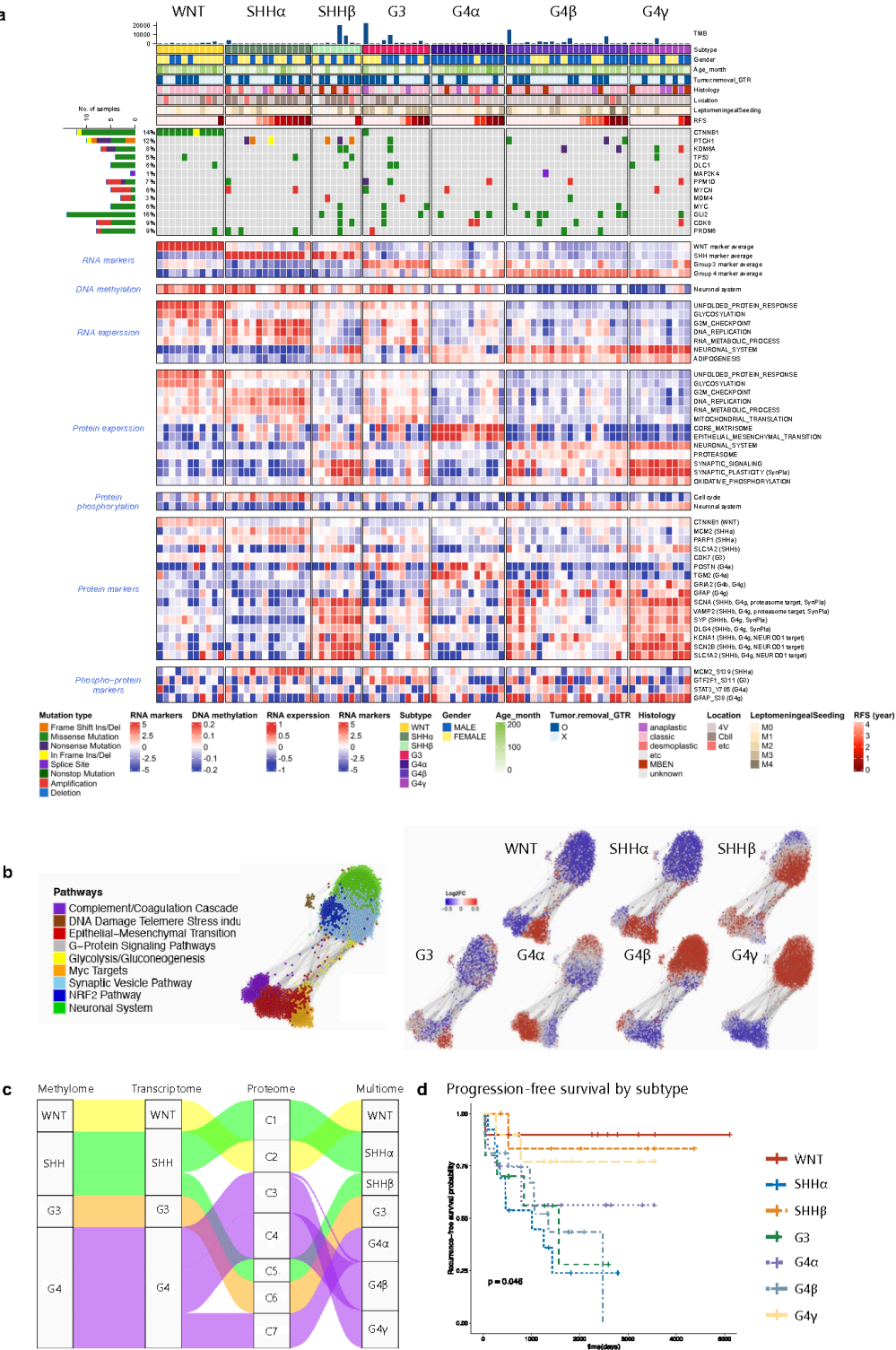
References

1. Ramaswamy V et al (2016) Risk stratification of childhood medulloblastoma in the molecular era: the current consensus. *Acta Neuropathol* 131:821–831. 10.1007/s00401-016-1569-6

2. Koschmann C, Bloom K, Upadhyaya S, Geyer JR, Leary SE (2016) Survival After Relapse of Medulloblastoma. *J Pediatr Hematol Oncol* 38:269–273. 10.1097/MPH.0000000000000547
3. Sabel M et al (2016) Relapse patterns and outcome after relapse in standard risk medulloblastoma: a report from the HIT-SIOP-PNET4 study. *J Neurooncol* 129:515–524. 10.1007/s11060-016-2202-1
4. Archer TC, Mahoney EL, Pomeroy SL (2017) Medulloblastoma: Molecular Classification-Based Personal Therapeutics. *Neurotherapeutics* 14:265–273. 10.1007/s13311-017-0526-y
5. Northcott PA et al (2017) The whole-genome landscape of medulloblastoma subtypes. *Nature* 547:311–317. 10.1038/nature22973
6. Cavalli FMG et al (2017) Intertumoral Heterogeneity within Medulloblastoma Subgroups. *Cancer Cell* 31, 737–754 e736. 10.1016/j.ccell.2017.05.005
7. Petralia F et al (1931) Integrated Proteogenomic Characterization across Major Histological Types of Pediatric Brain Cancer. *Cell* 183, 1962–1985 e 10.1016/j.cell.2020.10.044 (2020)
8. Hovestadt V et al (2019) Resolving medulloblastoma cellular architecture by single-cell genomics. *Nature* 572:74–79. 10.1038/s41586-019-1434-6
9. Archer TC et al (2018) Proteomics, Post-translational Modifications, and Integrative Analyses Reveal Molecular Heterogeneity within Medulloblastoma Subgroups. *Cancer Cell* 34, 396–410 e398. 10.1016/j.ccell.2018.08.004
10. Kool M 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. 10.1007/s00401-012-0958-8
11. Leal LF et al (2018) Reproducibility of the NanoString 22-gene molecular subgroup assay for improved prognostic prediction of medulloblastoma. *Neuropathology* 38:475–483. 10.1111/neup.12508
12. Smith KS et al (2022) Unified rhombic lip origins of group 3 and group 4 medulloblastoma. *Nature* 609:1012–1020. 10.1038/s41586-022-05208-9
13. Stromecki M et al (2018) Characterization of a novel OTX2-driven stem cell program in Group 3 and Group 4 medulloblastoma. *Mol Oncol* 12:495–513. 10.1002/1878-0261.12177
14. Yin WC et al (2019) Dual Regulatory Functions of SUFU and Targetome of GLI2 in SHH Subgroup Medulloblastoma. *Dev Cell* 48, 167–183 e165. 10.1016/j.devcel.2018.11.015
15. Forget A et al (2018) Aberrant ERBB4-SRC Signaling as a Hallmark of Group 4 Medulloblastoma Revealed by Integrative Phosphoproteomic Profiling. *Cancer Cell* 34, 379–395 e377. 10.1016/j.ccell.2018.08.002

Figures

Park et al. Fig. 1 Multiome profile



and protein phosphorylation levels, indicating subtype-specific molecular signatures and pathway activations. b. Visualization of protein function and expression patterns for each medulloblastoma subtype using a protein-protein correlation network. The networks are color-coded by distinct pathways, the activity of which are higher in each specific molecular subgroup. c. Sankey diagram illustrating the migration of samples through clustering of different molecular layers. d. Kaplan-Meier curve depicting progression-free survival (PFS) for patients stratified by medulloblastoma subtype ($p = 0.046$ by stratified log-rank test).

Park et al. Fig. 2 Subtypes associated NeuroD1-dependant differentiation

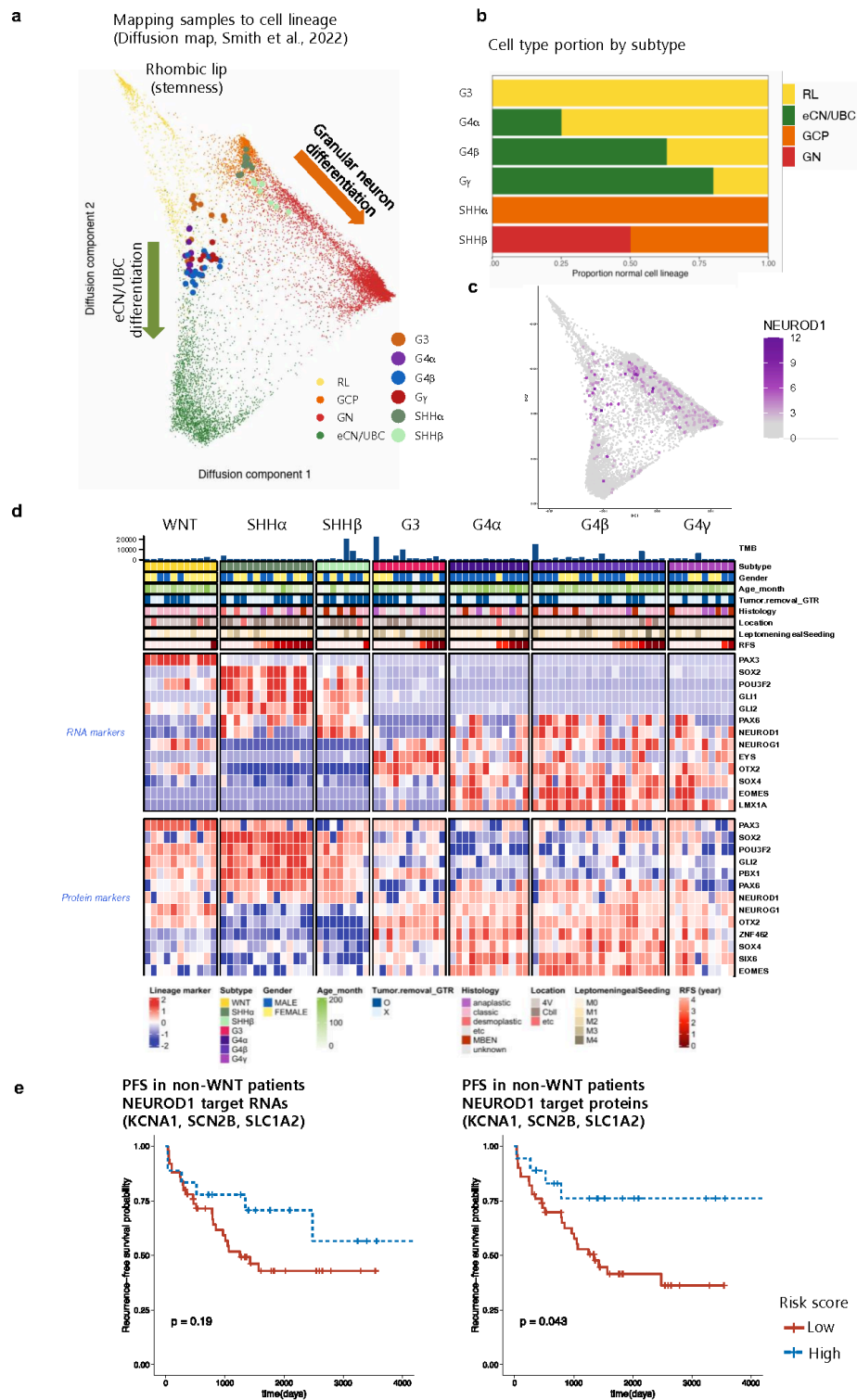


Figure 2

Medulloblastoma subtypes in the context of normal cerebellar development and differentiation. a. Mapping of medulloblastoma subtypes on scRNA-seq data of normal cerebellar cellular differentiation using a diffusion map. The plot shows different medulloblastoma subtypes with respect to normal neuronal differentiation from rhombic lip (stemness) to granular neuron / unipolar brush cell differentiation. b. Proportion of neuronal cell types within each medulloblastoma subtype. The bar chart visualizes the composition of neuronal cell lineages (RL, GCP, eCN/UBC, GN) across different medulloblastoma subtypes. c. Expression patterns of a representative neuronal differentiation marker (NEUROD1) on Diffusion map. d. Expression patterns of representative lineage markers across primary medulloblastoma subtypes. e. PFS difference by NEUROD1 target RNA and protein expression in non-WNT medulloblastoma patients ($p = 0.043$ by protein markers, $p = 0.19$ by RNA markers by log-rank test).

Park et al. Fig. 3 prognosis of SHH subgroup

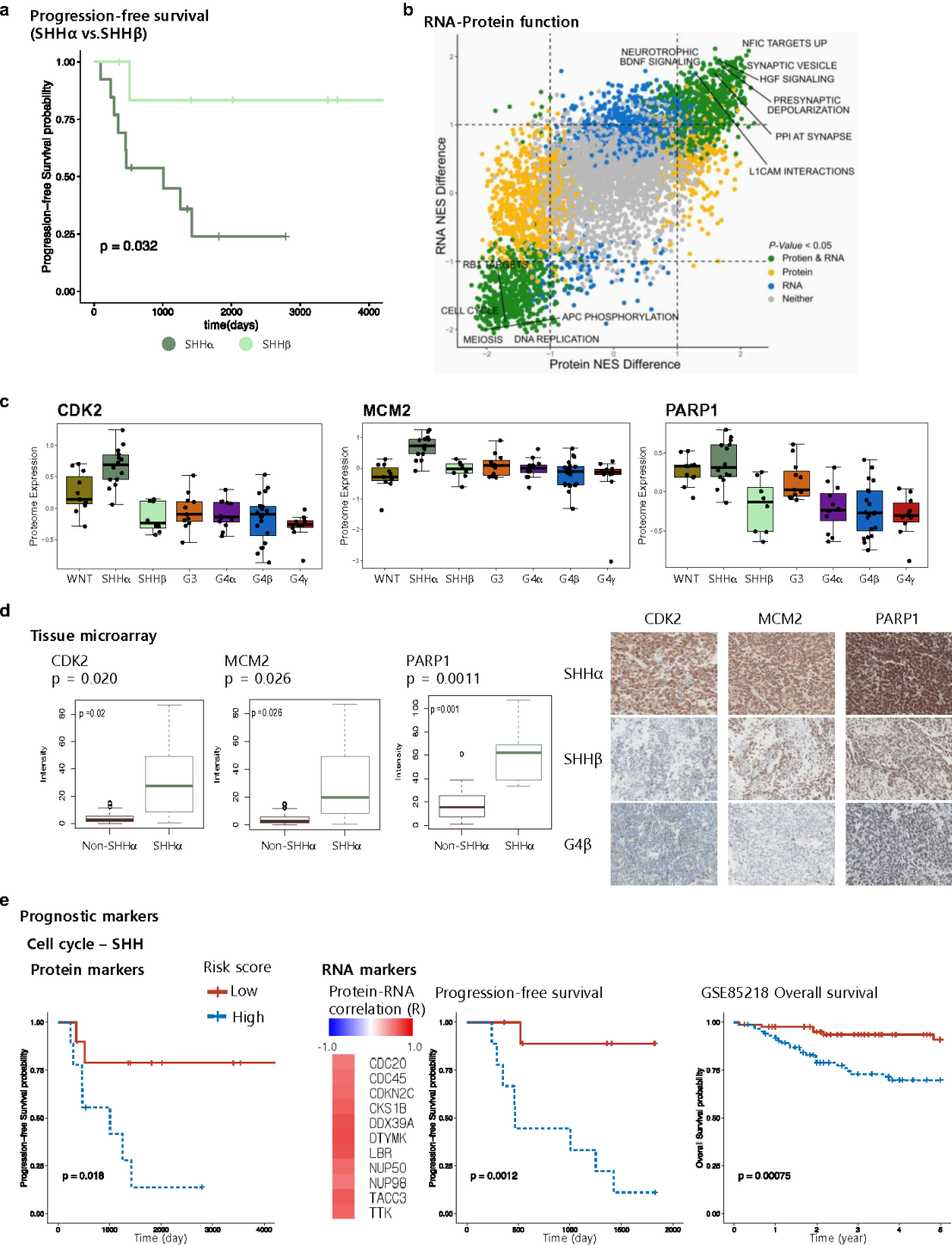


Figure 3

Comparative analysis of SHH α and SHH β subtypes in RNA and protein expression (Prognosis). a. PFS difference between SHH α and SHH β subtypes. b. Selection of RNA and Protein functions enriched in SHH α and SHH β subtypes (normalized enrichment score (NES)). c. Expression levels of representative proteins in SHH α subtype. d. Validation of SHH α -specific protein expression using tissue microarray

(TMA). e. Key prognostic protein and RNA markers exhibiting high correlation, identified using the Cox proportional hazards model.

Park et al. Fig. 4 Therapeutics of SHH

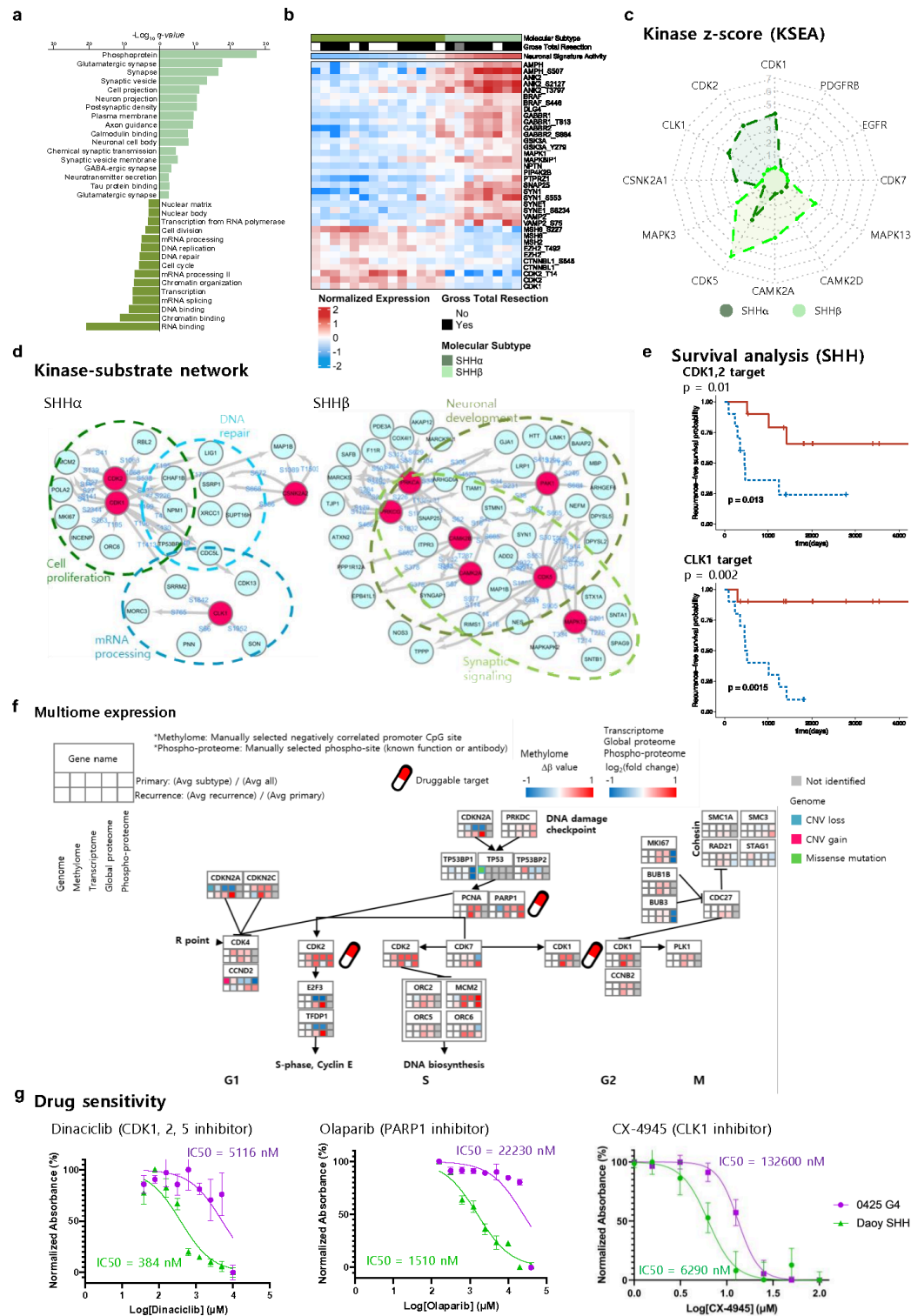


Figure 4

Difference between SHHa and SHHb subtypes in protein phosphorylation (Therapeutics). a. Enriched function of phosphorylated proteins in SHHa and SHHb subtypes. b. Representative phosphorylated proteins in SHHa and SHHb subtypes. c. Different kinase activity between SHHa and SHHb (KSEA kinase

z-score). d. Kinase-substrate network in SHHaand SHHb subtypes. e. Associations between PFS and the phosphorylation of subtype-specific kinase target proteins. (Coxph model, CDK1, 2 targets: p = 0.01, CLK1 targets: p = 0.002 by log-rank test). f. Multiome profile on cell cycle pathway in SHHa subtype. g. Comparison of drug sensitivity between SHH and Group 4 cell lines (Green: SHH (Daoy) cell line, purple: Group 4 (0425) cell line).

Park et al. Fig. 5. G4 subtypes (global and phospho-proteome)

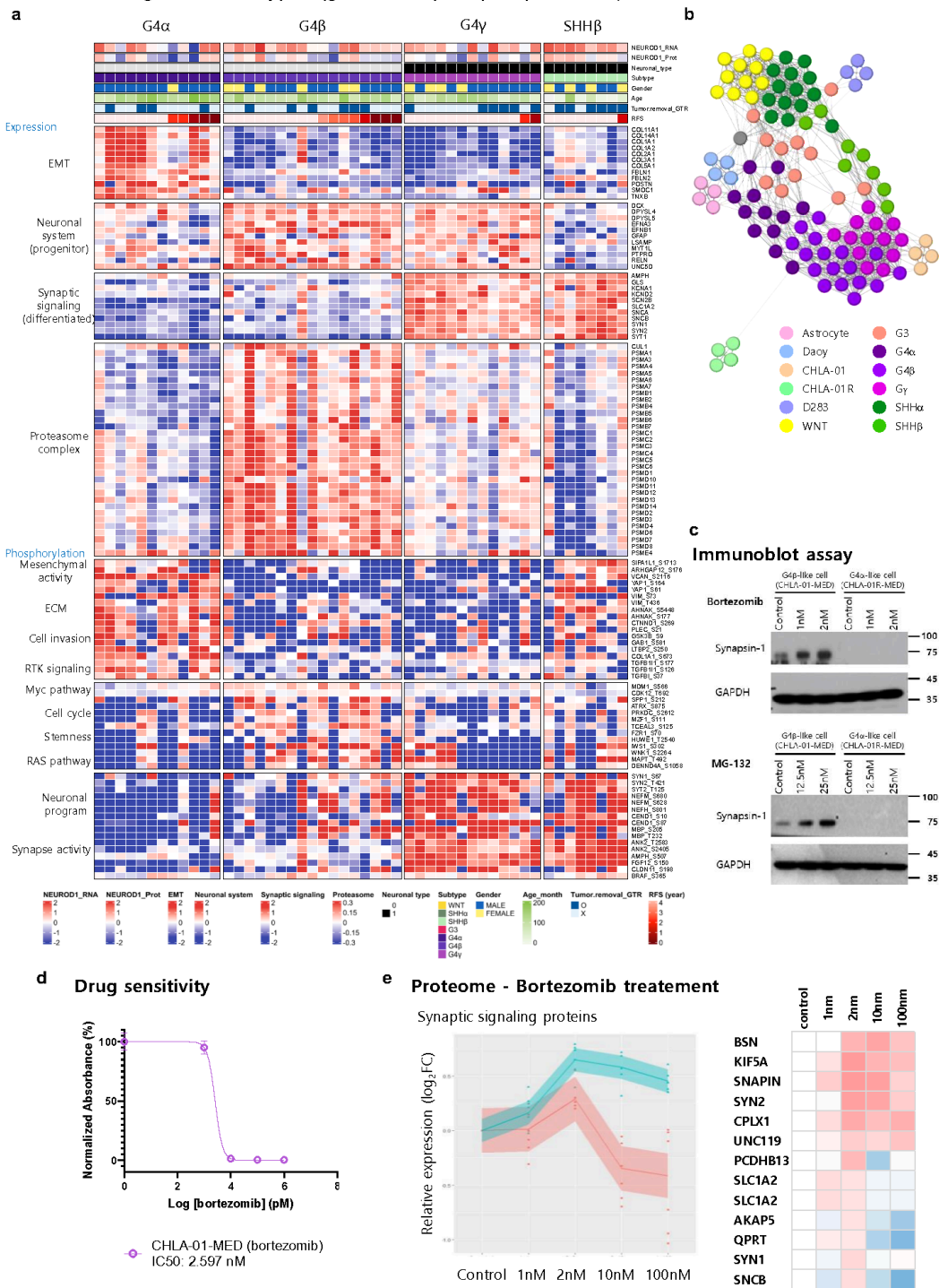


Figure 5

Proteomic features of Group4 subtypes. a. Representative protein expression and phosphorylation patterns of G4a, G4b, G4gand SHHb. b. Identification of subtypes within medulloblastoma cell lines in sample-sample correlation network using proteome data. c. Recovery of a neuronal differentiation marker (Synapsin-1, Syn1) by proteasome inhibitor treatment (immunoblot assay). d. Proteasome inhibitor sensitivity of a Group 4 cell line (CHLA-01-MED cell, G4b-like). e. Expression patterns of representative synaptic signaling proteins by dose-dependent Bortezomib treatment (LC-MS/MS)

Park et al. Fig. 6. Recurrence

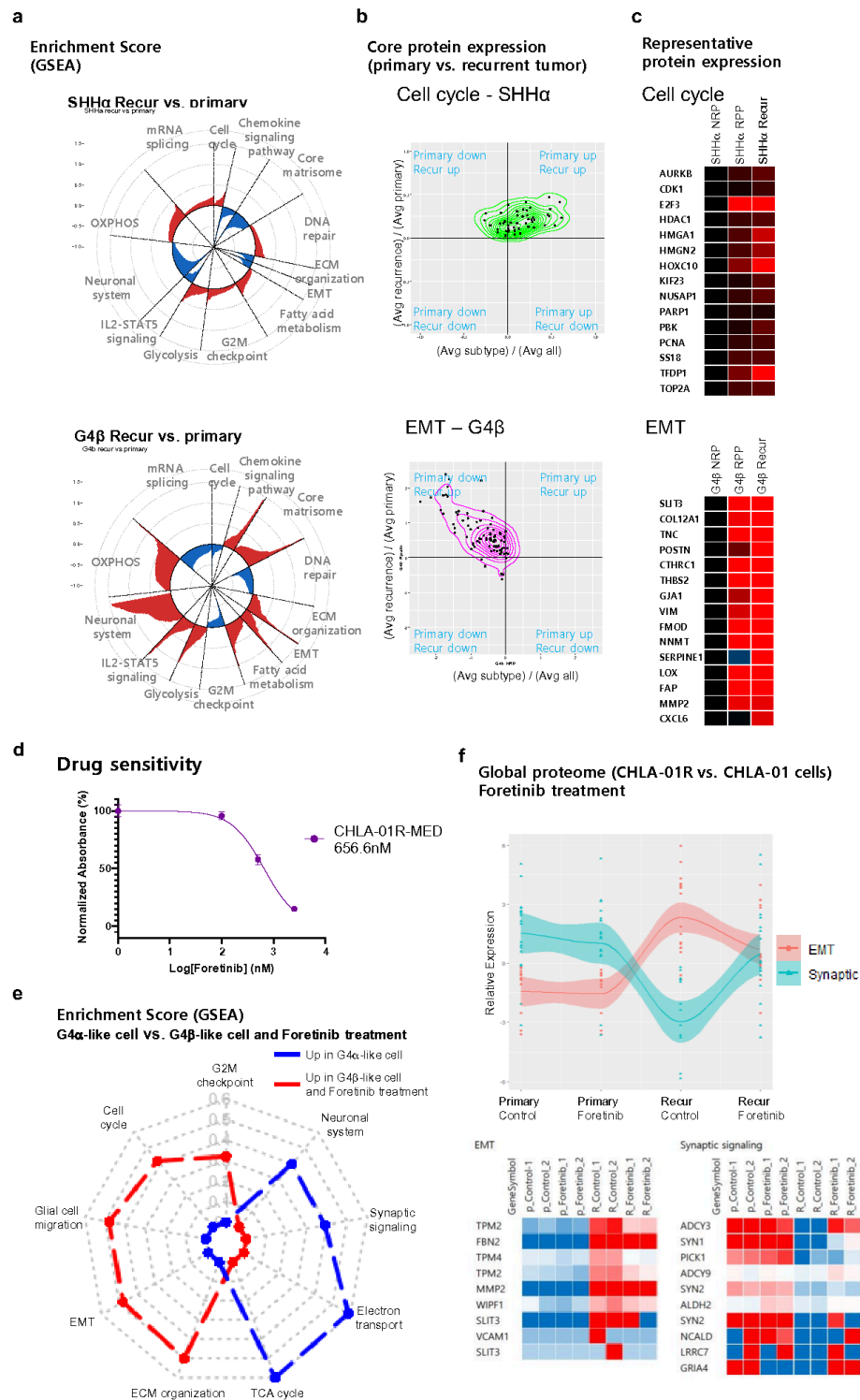


Figure 6

Subtype-specific protein changes during recurrence of medulloblastoma. a. Changes (GSEA enrichment score) in enriched protein functions during recurrence. b. Distribution of functional core protein expression (x-axis: change in primary tumors, y-axis: change during recurrence). c. Expression patterns of representative core proteins during recurrence. (NRP: Nonrecurrent primary tumor, RPP: Recurrent-potent primary tumor, Recur: Recurred tumor). d. MET inhibitor sensitivity of a Group 4 cell line (CHLA-01R-MED cell, G4a-like, recurrent). e. Enriched functions of medulloblastoma cell lines with MET inhibitor treatment. f. Protein function changes by MET inhibitor treatment (p: CHLA-01 cell (primary), R: CHLA-01R cell (recurrence)).

Park et al. Fig. 7. Summary

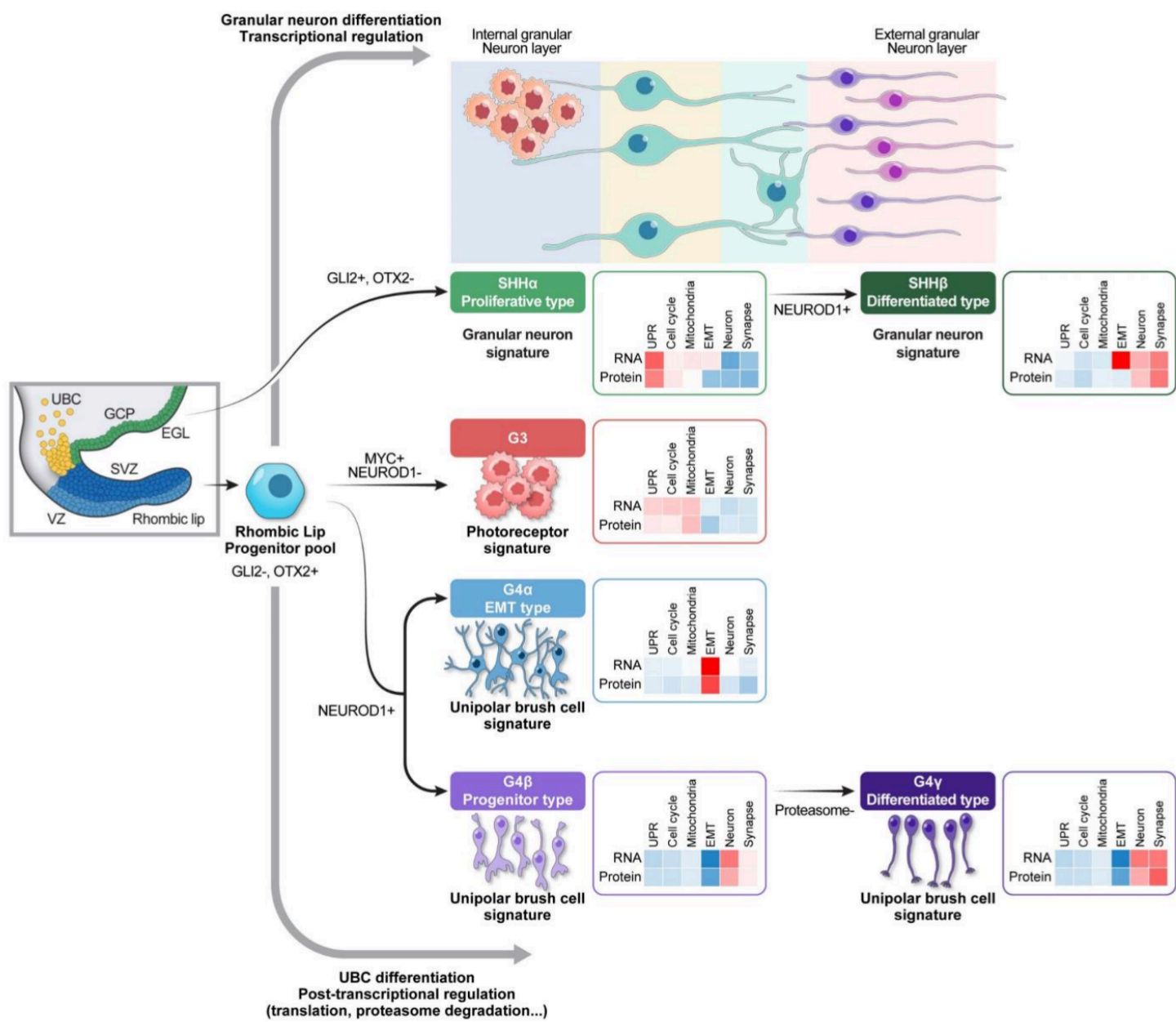


Figure 7

Summary of medulloblastoma progression. This Figure underscores the diverse cellular origins of medulloblastoma subtypes, detailing their specific differentiation pathways and regulatory mechanisms, which are crucial for targeted therapeutic approaches. (SVZ: subventricular zone, VZ: ventricular zone, GCP: granule cell progenitor, UBC: unipolar brush cell, EGL: external granule layer)

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [ExtendedDataFig.pdf](#)