Imaging Clusters of Pediatric Low-Grade Glioma are Associated with Distinct Molecular Characteristics

Anahita Fathi Kazerooni, Adam Kraya, Komal S. Rathi, Meen Chul Kim, Varun Kesherwani, Ryan Corbett, Arastoo Vossough, Nastaran Khalili, Deep Gandhi, Neda Khalili Ariana M. Familiar, Run Jin, Xiaoyan Huang, Yuankun Zhu, Alex Sickler, Matthew R. Lueder, Saksham Phul, Phillip B. Storm, Jeffrey B. Ware, Jessica B. Foster, Sabine Mueller, Jo Lynne Rokita, Michael J. Fisher, Adam C. Resnick and Ali Nabavizadeh

AJNR Am J Neuroradiol published online 14 February 2025
http://www.ajnr.org/content/early/2025/02/14/ajnr.A8699

Published February 14, 2025 as 10.3174/ajnr.A8699

ORIGINAL RESEARCH

Imaging Clusters of Pediatric Low-Grade Glioma are Associated with Distinct Molecular Characteristics

Anahita Fathi Kazerooni*, Adam Kraya*, Komal S. Rathi, Meen Chul Kim, Varun Kesherwani, Ryan Corbett, Arastoo Vossough, Nastaran Khalili, Deep Gandhi, Neda Khalili Ariana M. Familiar, Run Jin, Xiaoyan Huang, Yuankun Zhu, Alex Sickler, Matthew R. Lueder, Saksham Phul, Phillip B. Storm, Jeffrey B. Ware, Jessica B. Foster, Sabine Mueller, Jo Lynne Rokita, Michael J. Fisher, Adam C. Resnick, Ali Nabavizadeh

ABSTRACT

BACKGROUND AND PURPOSE: Cancers show heterogeneity at various levels, from genome to radiological imaging. This study aimed to explore the interplay between genomic, transcriptomic, and radiophenotypic data in pediatric low-grade glioma (pLGG), the most common group of brain tumors in children.

MATERIALS AND METHODS: We analyzed data from 201 pLGG patients in the Children's Brain Tumor Network (CBTN), using principal component analysis and K-Means clustering on 881 radiomic features, along with clinical variables (age, sex, tumor location), to identify imaging clusters and examine their association with 2021 WHO pLGG classifications. To determine the transcriptome pathways linked to imaging clusters, we employed a supervised machine learning model with elastic net logistic regression based on the pathways identified through gene set enrichment and gene co-expression network analyses.

RESULTS: Three imaging clusters with distinct radiomic characteristics were identified. *BRAF V600E* mutations were primarily found in imaging cluster 3, while *KIAA1549::BRAF* fusion occurred in subtype 1. The model's predictive accuracy (AUC) was 0.77 for subtype 1, 0.78 for subtype 2, and 0.70 for subtype 3. Each imaging cluster exhibited unique molecular mechanisms: subtype 1 was linked to oxidative phosphorylation, *PDGFRB*, and interleukin signaling, whereas subtype 3 was associated with histone acetylation and DNA methylation pathways, related to *BRAF V600E* pLGGs.

CONCLUSIONS: Our radiogenomics study indicates that the intrinsic molecular characteristics of tumors correlate with distinct imaging subgroups in pLGG, paving the way for future multi-modal investigations that may enhance understanding of disease progression and targetability.

ABBREVIATIONS: WHO = World Health Organization; CBTN =Children's Brain Tumor Network; pLGG = pediatric Low-Grade Glioma; EFS = Event-Free Survival; PC = Principal Component; CNS = Central Nervous System.

Received month day, year; accepted after revision month day, year.

From the Division of Neurosurgery (A.F.K., A.K., K.R., M.C.K., V.K., N.K., A.M.F., R.J., N.K., D.G., X.H., Y.Z., A.S., M.R.L., S.P., P.B.S., A.C.R., A.N.), Centre for Data-Driven Discovery in Biomedicine (D^{3}b) (A.F.K., A.K., K.R., M.C.K., V.K., N.K., A.M.F., R.J., N.K., D.G., X.H., Y.Z., A.S., M.R.L., S.P., P.B.S., A.C.R., A.N.), Department of Radiology (A.V.), Department of Pathology and Laboratory Medicine (M.R.L.), Division of Oncology (J.B.F., M.J.F.), The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, USA; Department of Neurosurgery (A.F.K., P.B.S., A.C.R.), Department of Radiology (A.N., J.B.W.), Department of Pediatrics (J.B.F.), University of Pennsylvania, Philadelphia, Pennsylvania, USA; Department of Pediatrics (R.C., J.L.R.), The George Washington University, Washington, DC, USA; Department of Neurology and Pediatrics (S.M.), University of California San Francisco, San Francisco, CA, USA.

The authors declare no conflicts of interest related to the content of this article.

Please address correspondence to Anahita Fathi Kazerooni, PhD, or Ali Nabavizadeh, MD, Center for Data-Driven Discovery in Biomedicine (D^{3}b), The Children's Hospital of Philadelphia, 734 Schuylkill Ave, Philadelphia, PA 19146; anahitaf@upenn.edu, Ali.Nabavizadeh@pennmedicine.upenn.edu.

SUMMARY SECTION

PREVIOUS LITERATURE: Radiogenomic studies on pediatric low-grade glioma often predict single genomic alterations, as outlined by WHO 2021 CNS 5. However, as molecular subtypes are increasingly incorporated into diagnostic procedures, predicting these alterations alone offers limited value for individualized treatment planning. In neuroscience, unsupervised clustering based on neuroimaging data is commonly used to identify group-level clinical and molecular traits and therapeutic targets. We propose applying this approach to pediatric low-grade glioma to identify molecular pathway groups associated

with specific radiographic phenotypes. This method could uncover novel therapeutic targets, bridging radiographic and molecular insights to advance personalized treatment strategies.

KEY FINDINGS: We identified three imaging clusters: Cluster 1 showed tumor growth mechanisms and suppressed immune activity; Cluster 2 exhibited metabolic plasticity and glioma stem cell traits; Cluster 3 had upregulated histone acetylation/DNA methylation and rare *KIAA1549::BRAF* fusions linked to poorer prognosis.

KNOWLEDGE ADVANCEMENT: This is the first study to investigate genotype-radiophenotype associations in pediatric low-grade glioma using comprehensive machine learning analysis of WGS and RNA-seq data. Our findings demonstrate that grouping tumors based on treatment-naïve imaging and clinical characteristics can uncover potential therapeutic targets.

INTRODUCTION

Cancer is a complex biological system influenced by a tumor's evolutionary environmental forces. Tumor phenotypes are shaped by a combination of interconnected molecular events across the genome, transcriptome, epigenome, and proteome among other cellular contexts, which collectively converge to dysregulate cellular biological functions in ways that promote oncogenesis and cancer progression [1]. Exploring the relationships between radiophenotypes—tumor traits beyond conventional analysis quantified from radiological imaging that reflect characteristics at genomic and transcriptomic levels—can enhance our understanding of tumor dynamics and the variability in therapeutic responses across similar histologies [2].

Pediatric low-grade gliomas (pLGGs), the most common brain tumors in children, account for one-third of all pediatric brain cancers [3, 4]. While complete resection improves survival, deeply-seated or infiltrative tumors often necessitate chemotherapy post-partial resection, leading to a 10-year event-free survival (EFS) of around 50% [5, 6]. Treatments can also impact cognitive and neurological outcomes in longer term, affecting quality of life and survivorship [7].

PLGGs encompass a wide array of molecular subtypes, each associated with different prognoses and responses to treatment, underscoring the need for treatments targeted to specific subtypes. The introduction of targeted treatments, including *RAF* or *MEK* inhibitors, has expanded treatment options for patients with pLGG, paving the way for improved patient outcomes [8, 9]. However, understanding the biological and molecular basis of pLGG tumors beyond single gene mutation is critical for the success of these targeted therapies and may be helpful to address current challenges of tumor recurrence, rebound, and developing resistance in the era of molecularly targeted agents [8, 10].

Radiomics – a high throughput approach for extracting quantitative features from radiological images – offers the potential to provide non-invasive biomarkers for characterizing the molecular underpinnings of tumors [11, 12]. However, most existing studies on the radiogenomic analysis of pLGGs focus on predicting individual gene alterations, such as *KIAA1549::BRAF* fusion or *BRAF V600E* mutation [13-16]. As WHO CNS 5 classification recommends the inclusion of driver molecular alterations in standard diagnosis, predicting a single mutation or fusion using imaging alone may not be useful or sufficient to fully understand the tumor biology. Moreover, focusing on a single pathway or gene mutation limits our understanding of the complex interactions between different pathways that drive patient-specific responses to targeted treatments.

In our preliminary work, we explored the association between imaging clusters, referred to as imaging subtypes, derived through an unsupervised machine learning approach based on radiomic features, and specific gene mutations, including *BRAF V600E*, in pLGGs [17]. In the present radiogenomic study, we aim to leverage unsupervised clustering of imaging characteristics to identify distinct and relatively homogeneous subgroups or subtypes [18] within pLGGs. This approach seeks to elucidate how information from genomic and transcriptomic scales translates into radiophenotypic characteristics. We hypothesize that radiological patterns, quantified through radiomic features and analyzed using unsupervised machine learning, can uncover molecular aspects of tumors and identify potential treatment targets, extending beyond the WHO-recognized molecular subtypes.

MATERIALS AND METHODS

The overall workflow of the analysis is illustrated in FIG 1, outlining the key steps: (1) MRI preprocessing and radiomic feature extraction, (2) unsupervised clustering to identify imaging clusters, (3) transcriptomic and genomic analyses to identify gene and pathway alterations, and (4) comparison of molecular characteristics across imaging clusters. The methodology proposed by "The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: guidelines for reporting observational studies" was followed.

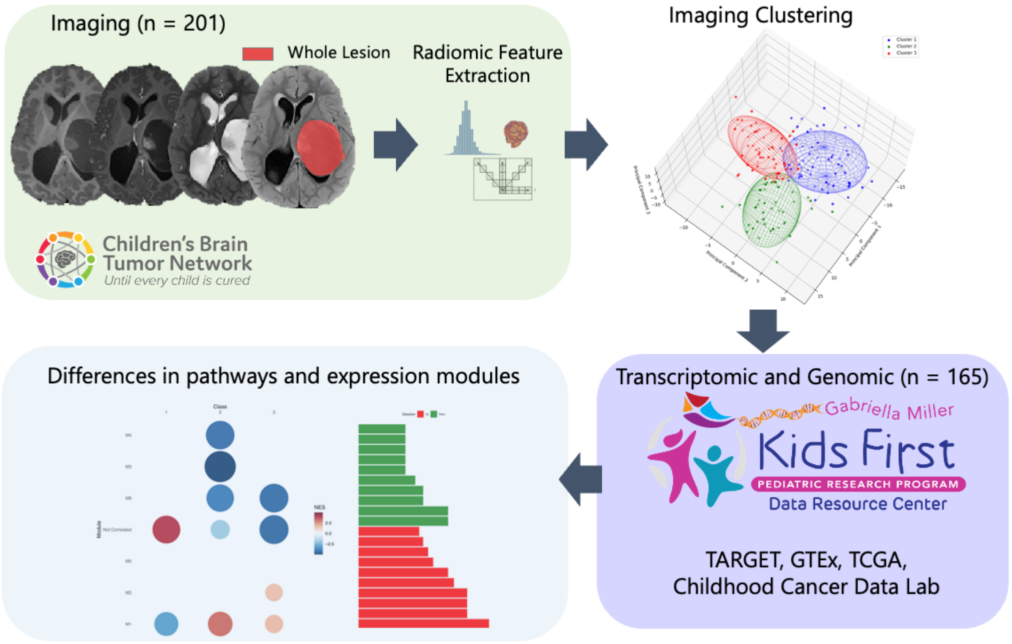


FIG 1. Graphical representation of the analysis steps.

Overview of the Data

In this study, we analyzed retrospectively-collected data from the Children's Brain Tumor Network (CBTN) [19] repository (cbtn.org), which includes specimens and longitudinal clinical and imaging data, facilitating the sharing of de-identified samples for research. The study complied with HIPAA guidelines and received IRB approval from the Children's Hospital of Philadelphia (CHOP) through the CBTN protocol, with informed consent obtained for patient enrollment.

We focused on patients histopathologically diagnosed with *de novo* pLGG from 2006 to 2018. Standard multiparametric MRI (mpMRI) sequences, including pre- and post-Gadolinium T1-weighted (T1w, T1w-Gd), T2-weighted (T2w), and T2 fluid-attenuated inversion recovery (T2-FLAIR), were collected from 258 pLGG subjects. Exclusion criteria included patients older than 18 years at the time of imaging, imaging conducted post-surgery or treatment, tumors outside the brain, leptomeningeal dissemination, missing MRI sequences, or low-quality imaging. After exclusions, 201 subjects with complete mpMRI sequences were included for analysis. Molecular data from 165 intersecting subjects were compiled from CBTN records and the Open Pediatric Cancer (OpenPedCan) project repository (version 15 data release; DOI: 10.5281/zenodo.6473912) [20]. Patient characteristics are summarized in Supplementary Table 1.

Imaging Data Analysis

Details on image preprocessing are provided in Supplementary Material SI1. Images underwent skull stripping [21] and intensity normalization to a scale of [0, 255] after removing outliers to facilitate radiomic feature extraction. We generated Whole Tumor (WT) masks by combining all four tumor components generated using our in-house automatic pediatric brain tumor segmentation tool (https://github.com/d3b-center/peds-brain-auto-seg-public) [21, 22], followed by manual revisions when necessary. An extensive array of radiomic features (n = 881), including metrics of shape, volume, intensity, and texture were extracted. The texture analyses included techniques such as gray-level co-occurrence matrix (GLCM), gray-level run-length matrix (GLRLM), gray-level size zone matrix (GLSZM), neighborhood gray tone difference matrix (NGTDM), local binary pattern (LBP), and Collage [23] features, using the WT masks on the multiparametric MRI images.

We randomly split the dataset into a discovery group (80% of the cohort; n = 160) for model training, and a replication group (20% of the cohort; n = 41) for testing, and conducted a pairwise Pearson's correlation analysis on the radiomic features within the discovery group, eliminating one feature from each pair that had a correlation coefficient (r) exceeding 0.90, reducing the number of features to 438. These features were normalized via z-scoring. Tumor location, a categorical variable, was normalized through count encoding, and age at diagnosis was standardized using z-scoring based on the mean and standard deviation of continuous features from the discovery group, which was then applied to normalize the features in the replication group. We produced histograms that display the distribution of signal intensities within specific sequences (x-axis) against their frequency (y-axis) across the entire dataset, providing insights into the characteristic distribution of these features among different pLGG imaging clusters.

Imaging-Based pLGG Clusters

We applied a multi-step clustering approach to a feature set comprising 441 variables, including 438 radiomic features, along with age, sex, and tumor location. To reduce the dimensionality of the feature space, we utilized principal component analysis (PCA) and retained 78 principal components (PCs), capturing over 95% of the total variance and preserving the essential information required for accurate modeling and analysis. To determine the optimal number of imaging-based pLGG clusters, we used the "elbow method", which analyzes the "within-cluster sum of squares" across the number of clusters ranging from 1 to 160, identifying the point of diminishing returns. To ensure robust cluster determination, we supplemented the elbow method with the "silhouette coefficient", calculated for cluster counts from 2 to 15. This metric, ranging from -1 to 1, validated the optimal cluster number by maximizing inter-cluster separation and intra-cluster cohesion.

After identifying the optimal number of clusters, we used the "Greedy K-means++" algorithm to cluster the PCs from the discovery cohort. This method, which is less sensitive to initial centroid selection, improves clustering reliability and stability. To mitigate any potential biases from random initialization and ensure reproducibility, the model was trained 1,000 times in a single run. The finalized model was then used to predict cluster membership for the replication cohort ($n = 41$), to evaluate generalizability.

Differential Gene and Pathway Expression Analyses

We analyzed clinical and RNA-Seq data from the CBTN [19] repository using the STAR-RSEM RNASeq pipeline of the OpenPBTA [20]. Details on gene set enrichment analysis (GSEA) can be found in Supplementary Material SI2. To identify pathway expressions that are most predictive of imaging cluster assignment, using the glmnet R package, we trained an ElasticNet logistic regression classifier using a one-versus-rest strategy for each imaging cluster, leveraging 70% of the low-grade glioma cohort in this study to train – validate the model and leaving 30% as an independent test set, and used 10-fold cross validation to tune the alpha and lambda model parameters. We selected the top 100 most variable differentially expressed/co-expressed pathways along with molecular subtype as described by Ryall et al [6, 20], the anatomical region of the tumor, age at diagnosis, reported sex, and race as predictors. We evaluated the area under the receiver operating curve (AUC-ROC) of the training – validation and independent testing sets to evaluate the degree of overfitting for each model and retained pathways with non-zero generalized linear model (GLM) coefficients for further evaluation.

Association Analysis between Imaging and pLGG WHO 2021 Classification

PLGG tumor entities described by the 2021 World Health Organization (WHO) classification of central nervous system (CNS) tumors [4] were compiled based on molecular and pathological information from the OpenPedCan v15 data repository (described before), which annotates pLGG with well-defined genomic lesions, including single nucleotide (SNVs), insertion-deletion (indel), and structural variants (SVs) occurring in MAPK pathway-related and non-MAPK pathway-related genes. Tumors were designated as 'Pediatric-type Diffuse LGG, NOS' if the patient's pathology report designated the tumor as such and if the patient had only a single molecular data modality available (DNA- or RNA-Seq), as both modalities are required for determining the molecular subtype of the tumor [6]. The association between imaging and WHO 2021 classifications was evaluated using a likelihood ratio test (LRT) between a saturated Poisson GLM and a nested Poisson model excluding the interaction term. A correspondence analysis was then performed on the Pearson residuals derived from the Poisson saturated model to visualize the associations between imaging and pLGG WHO CNS tumor classifications.

Survival Analysis

Kaplan-Meier analysis of EFS was performed to compare patient outcomes by cluster assignment. Cox proportional-hazards regression models were also developed, including covariates for the extent of tumor resection and the interaction between tumor molecular subtype and cluster assignment to identify predictors of survival. Loss to follow-up was treated as censored data, with EFS defined as the time from diagnosis to the last available follow-up. All survival analyses were performed using the *survival* R package (2020; https://CRAN.R-project.org/package=survival), and Kaplan-Meier survival curves were generated using the *survminer* R package (2021; https://rpkgs.datanovia.com/survminer/index.html).

Analysis of KIAA1549::BRAF Fusion Breakpoints

All high-confidence, in-frame *KIAA1549*::*BRAF* STAR-fusion and/or Arriba fusion calls in pLGG tumors were annotated with exon number in canonical transcripts (NM_001164665 for *KIAA1549* and NM_004333 for *BRAF*) using *biomaRt* and

GenomicRanges R packages [24-26]. Common breakpoints included those involving exons 15:09 (exon 15 in *KIAA1549* and exon 9 in *BRAF*), 16:09, 16:11, and 18:10 [6, 27]. All other breakpoint combinations were classified as rare/novel.

Data and Code Availability

The processed data and the codes utilized for the analysis of genomic, transcriptomic, and radiomic data in conjunction with patient clinical characteristics is accessible at https://github.com/d3b-center/pLGG_imaging_clustering_genomic. All image processing tools employed in this study are publicly available and free to use, including CaPTk (https://www.cbica.upenn.edu/captk) and the in-house automated tumor segmentation model (https://github.com/d3b-center/peds-brain-auto-seg-public).

RESULTS

Characteristics of Imaging Clusters

Our clustering model stratified patients into three imaging clusters (FIG 2A, Supplementary FIG 1) and was applied to the replication set. Histograms (Supplementary FIG 2), presenting variations in imaging characteristics across the three imaging clusters, showed cluster 2 had moderate contrast enhancement based on the location of the peak of histogram falling around intermediate intensity values. This cluster also shows higher peak of T2-FLAIR intensity values within peritumoral edema in the higher range of intensity values. Additionally, reduced T2 intensity values were observed in the non-enhancing tumor regions for imaging cluster 2 compared to the other clusters, which might suggest a denser tumor cell population.

Top radiomic features influencing PCs (FIG 2B) mainly included the texture features that reflect tumor heterogeneity, such as GLCM contrast and homogeneity features and the coefficient of variation and skewness from the first-order histogram feature family. Additionally, a morphologic feature was also identified as a key contributor to the composition of the PCs.

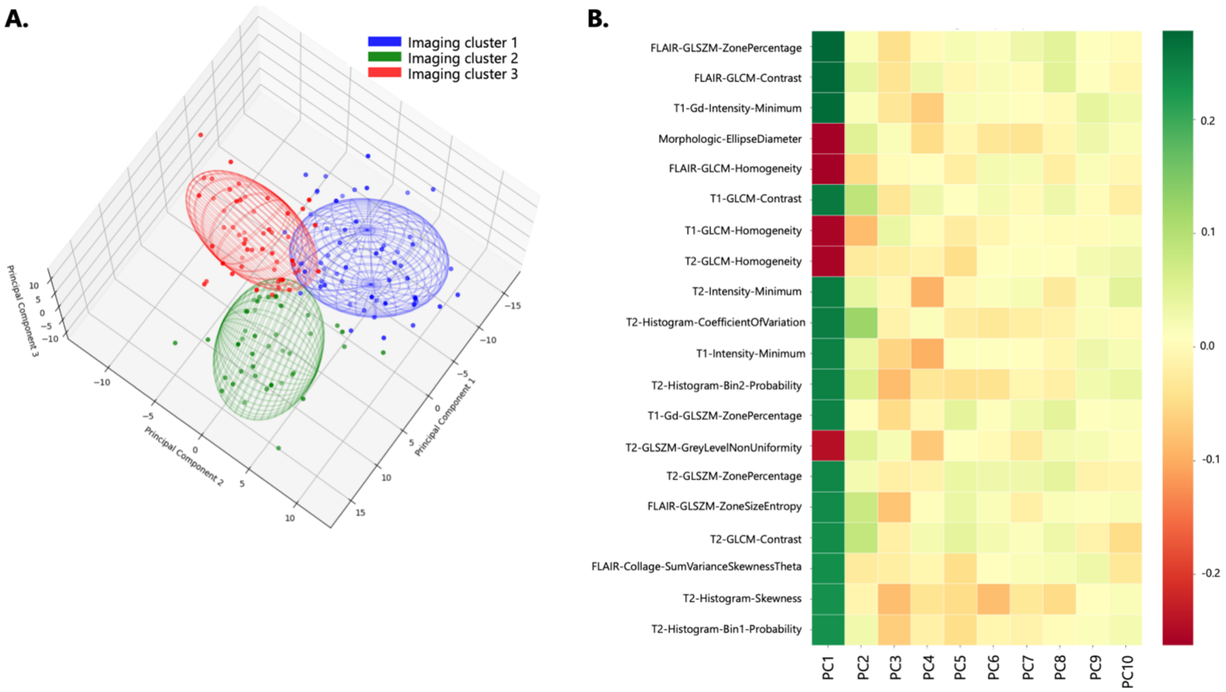


FIG 2. (A) Three-dimensional visualization of imaging clusters in the discovery set; (B) Heatmap plot of the radiomic features contributing to the top 10 principal components (PCs).

Relationship Between Imaging Clusters and WHO CNS Tumor Classifications

Correspondence analysis of Pearson residuals from a Poisson GLM revealed distinct associations between imaging clusters and the 2021 WHO CNS tumor classifications [4] (FIG 3A; Supplementary Table 2). Cluster 1 was enriched with pilocytic astrocytomas, predominantly cerebellar, as observed in other studies [9, 28], with *KIAA1549::BRAF* fusions (IRR = 4.67, p<0.001).

Cluster 2 was over-represented with *MAPK*-altered diffuse low-grade gliomas (IRR = 4.67, p<0.001), harboring SNV-indel mutations in either *KRAS*, *NRAS*, *HRAS*, *MAP2K1/2*, *ARAF*, *RAF1*, or *BRAF*, or germline or somatic *NF1* variants. It was also enriched for gangliogliomas (IRR = 11.2, p = 0.022), but under-represented with pilocytic astrocytomas (IRR = 0.6, p = 0.034).

Cluster 3 was under-represented with pilocytic astrocytomas (IRR = 0.5, p<0.001) and enriched with *MAPK*-altered diffuse low-grade gliomas (IRR = 2.85, p<0.001) and pediatric-type diffuse low-grade gliomas, NOS. Cluster 3 tumors, including *BRAF V600E* mutations, were distributed across various brain regions (FIG 3B).

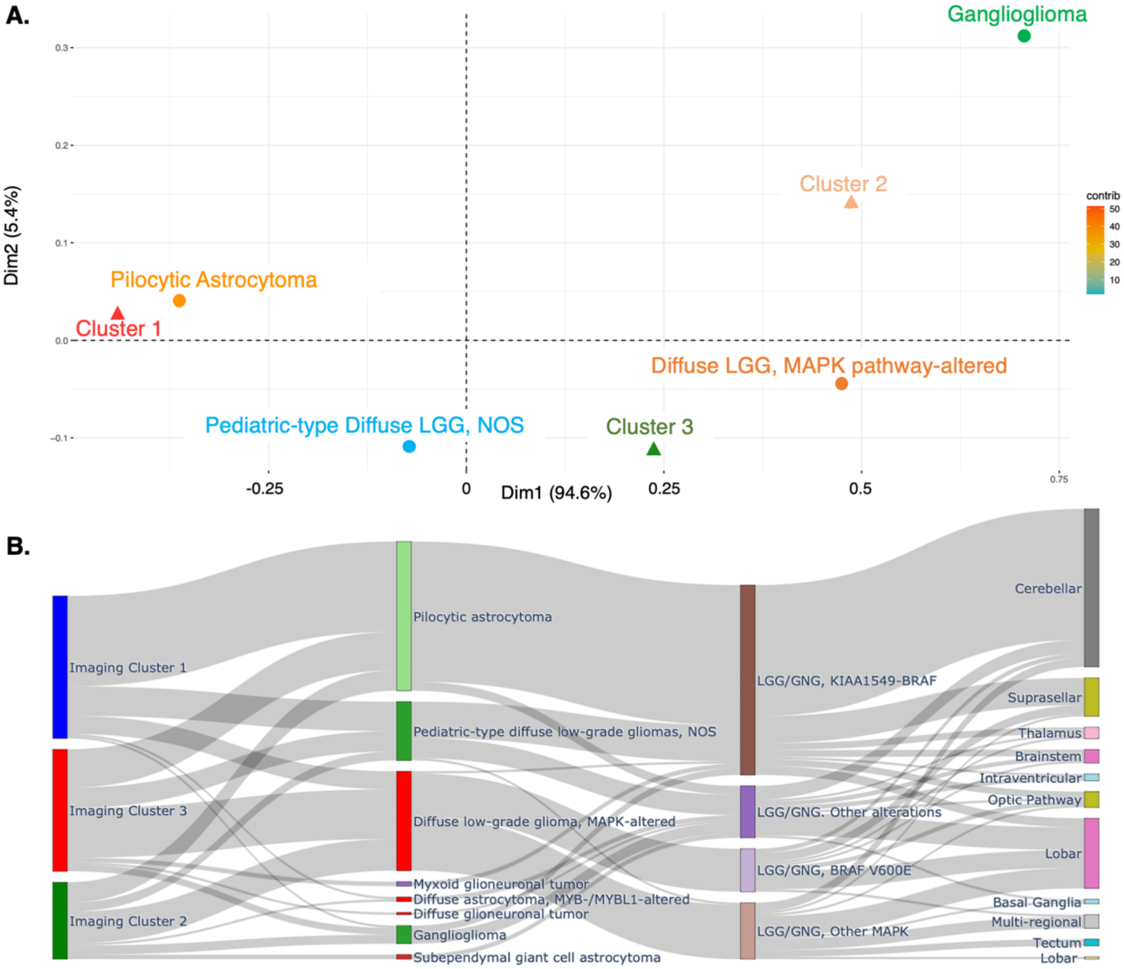


FIG 3. (A) correspondence analysis plot of Pearson residuals depicting the association between imaging clusters and pLGG 2021 WHO CNS Tumor Classifications. (B) Alluvial plot indicating associations of imaging clusters with pLGG 2021 WHO CNS Tumor Classifications, pLGG molecular subtypes by Ryall et. al [6], and tumor locations.

Transcriptomic Pathway Analysis

Gene set enrichment analysis (GSEA) (FIGs 4A-C) revealed distinct pathway activities across clusters. Cluster 1, in comparison with Cluster 2, exhibited lower fatty acid metabolism, critical for glioma growth [29, 30], and rhodopsin-like *GPCR* signaling [31] but higher oxidative phosphorylation and *PDGFRB*-related pathways [5, 32].

Cluster 1 in comparison with Cluster 3, showed reduced rhodopsin-like receptor and *FGFR1* signaling, frequently mutated in midline low grade gliomas [33, 34], along with fewer *FGFR1* mutations. Pathways elevated in Cluster 1 included oxidative phosphorylation and immune regulatory pathways (complement cascade, MHC class I antigen presentation).

Cluster 2 in comparison with Cluster 3, showed higher expression in adhesion-related and extracellular matrix remodeling pathways, suggesting increased migratory potential in these tumors.

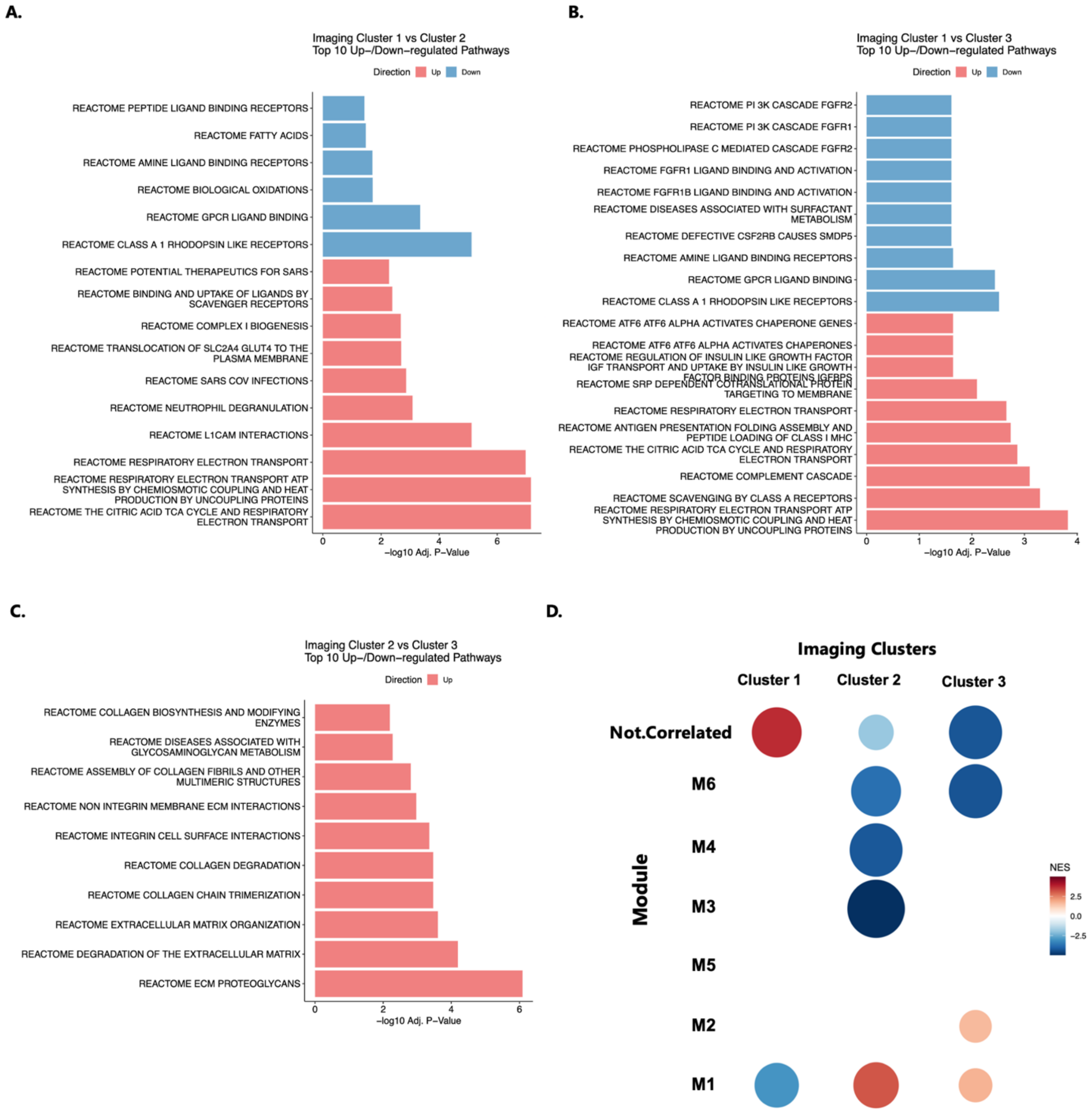


FIG 4. Molecular pathways associated with imaging clusters: (A-C) Top GSEA pathways differentially up-regulated or down-regulated between pairs of imaging clusters; (D) matrix of Pearson's residuals for co-expression modules (defined by CEMITool) across the three imaging clusters.

Gene Co-Expression Module Analysis

Gene co-expression analysis (CEMITool) identified seven differentially expressed modules across imaging clusters (Supplementary Material SI2; FIG 4D). Module M1 showed lower expression in Cluster 1 for rhodopsin-like GPCRs, potassium channel signaling, GABA signaling, and synaptic neurotransmission. Module M2 indicated higher activity in Cluster 3 for histone acetylation and DNA methylation pathways. Module M3 showed reduced activity in Cluster 2 for potassium signaling and glucose homeostasis pathways. Module M4 demonstrated lower levels of interleukin (IL-4, IL-6, IL-10, IL-13) and PD-1 signaling in Cluster 2. Module M6 showed reduced extracellular matrix remodeling and integrin-related pathway expression in Clusters 2 and 3, with Cluster 3 showing the lowest activity. Non-Correlated genes were enriched in tumor-promoting interleukin pathways (e.g., IL-4, IL-10, IL-13 [35-38]), particularly in Cluster 1. Module M5 showed no significant differences or biological enrichment across clusters.

Machine Learning-Based Cluster Classification Based on Molecular Characteristics

Leveraging a machine learning-based approach to compare differentially expressed/co-expressed signaling pathways, molecular subtypes, and clinical variables across imaging clusters achieved the following predictive performance for imaging clusters (Supplementary FIG 3): Cluster 1, AUC-ROC of 0.91 (training-validation, CI: 0.86-0.96) | 0.78 (testing, CI: 0.654-0.91), with a balanced accuracy of (0.83, training); Cluster 2, AUC-ROC of 1 (training-validation, CI: 0.99-1) | 0.84 (testing, CI: 0.7-0.98), with a balanced accuracy of 0.81 (training); Cluster 3, AUC-ROC of 0.84 (training-validation, CI: 0.76-0.91) | 0.69 (testing, CI: 0.63 - 0.74), with a balanced accuracy of 0.68 (training).

GLM coefficients for each classifier are shown in FIGs 5A-C, highlighting key features influencing classification. In Cluster 1 (FIG 5A), tumor anatomical location and stem-like signaling via *Notch2* were positively associated, and deleterious *BRAF* mutation-driven subtypes and oxidative stress-induced senescence were negatively associated.

In Cluster 2 (FIG 5B), caspase activation and oxidative stress-induced senescence were positively associated, and deleterious *BRAF* alterations, stem-like *Notch2* signaling, and cholesterol biosynthesis were negatively associated.

In Cluster 3 (FIG 5C), AMPA receptor signaling and cholesterol biosynthesis were positively associated, and innate immune signaling was negatively associated.

When generating three additional independent training-testing splits of the data, we observed reproducible sensitivity from our predictive algorithm (Supplementary FIG 4).

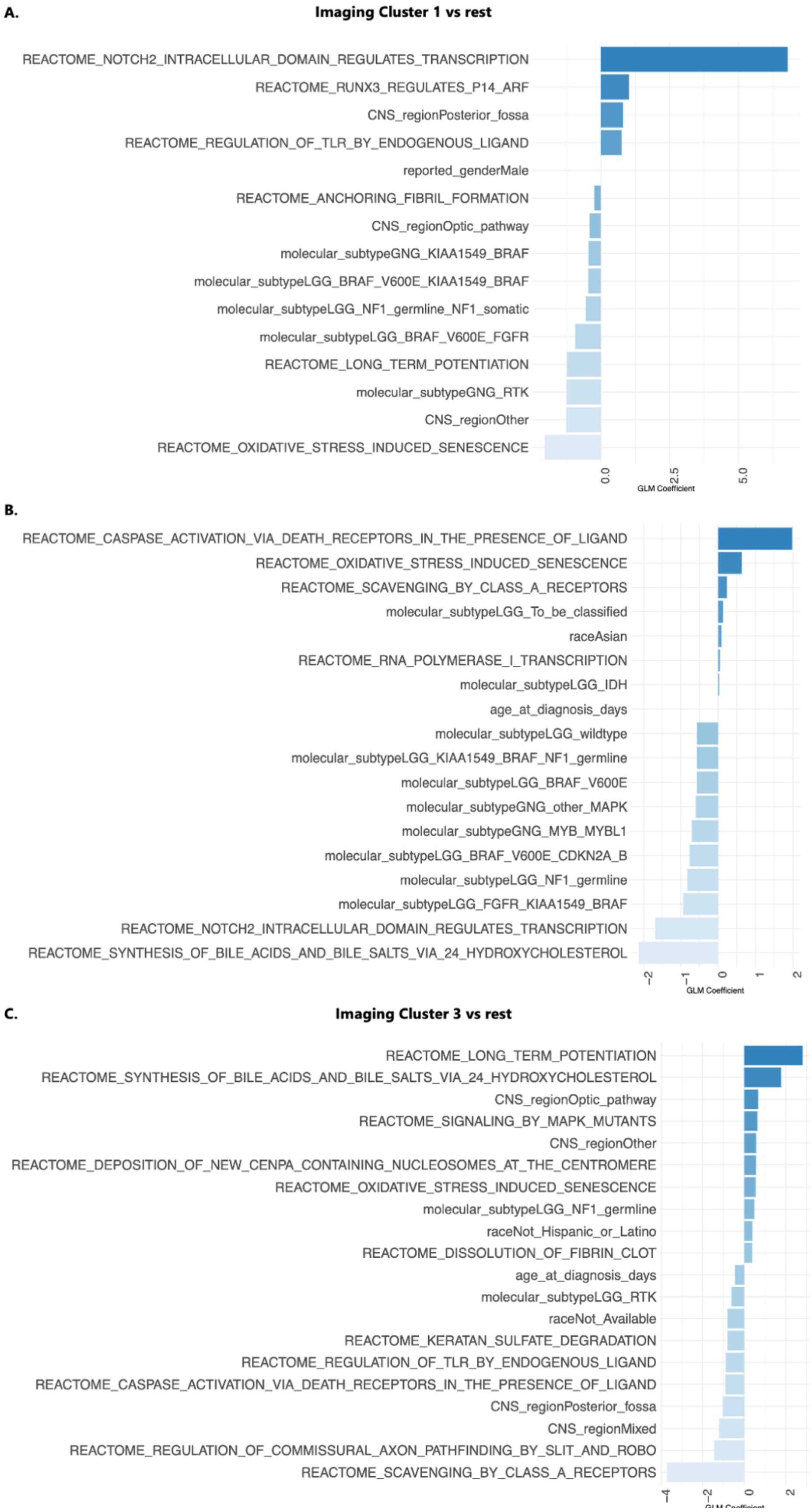

A.
Imaging Cluster 1 vs rest
REACTOME_NOTCH2_INTRACELLULAR_DOMAIN_REGULATES_TRANSCRIPTION
REACTOME_RUNX3_REGULATES_P14_ARF
CNS_regionPosterior_fossa
REACTOME_REGULATION_OF_TLR_BY_ENDOGENOUS_LIGAND
reported_genderMale
REACTOME_ANCHORING_FIBRIL_FORMATION
CNS_regionOptic_pathway
molecular_subtypeGNG_KIAA1549_BRAF
molecular_subtypeLGG_BRAF_V600E_KIAA1549_BRAF
molecular_subtypeLGG_NF1_germline_NF1_somatic
molecular_subtypeLGG_BRAF_V600E_FGFR
REACTOME_LONG_TERM_POTENTIATION
molecular_subtypeGNG_RTK
CNS_regionOther
REACTOME_OXIDATIVE_STRESS_INDUCED_SENESCENCE
0.0
2.5
5.0
GLM Coefficient
B.
REACTOME_CASPASE_ACTIVATION_VIA_DEATH_RECEPTORS_IN_THE_PRESENCE_OF_LIGAND
REACTOME_OXIDATIVE_STRESS_INDUCED_SENESCENCE
REACTOME_SCAVENGING_BY_CLASS_A_RECEPTORS
molecular_subtypeLGG_To_be_classified
raceAsian
REACTOME_RNA_POLYMERASE_I_TRANSCRIPTION
molecular_subtypeLGG_IDH
age_at_diagnosis_days
molecular_subtypeLGG_wildtype
molecular_subtypeLGG_KIAA1549_BRAF_NF1_germline
molecular_subtypeLGG_BRAF_V600E
molecular_subtypeGNG_other_MAPK
molecular_subtypeGNG_MYB_MYBL1
molecular_subtypeLGG_BRAF_V600E_CDKN2A_B
molecular_subtypeLGG_NF1_germline
molecular_subtypeLGG_FGFR_KIAA1549_BRAF
REACTOME_NOTCH2_INTRACELLULAR_DOMAIN_REGULATES_TRANSCRIPTION
REACTOME_SYNTHESIS_OF_BILE_ACIDS_AND_BILE_SALTS_VIA_24_HYDROXYCHOLESTEROL
−2
−1
0
1
2
GLM Coefficient
C.
Imaging Cluster 3 vs rest
REACTOME_LONG_TERM_POTENTIATION
REACTOME_SYNTHESIS_OF_BILE_ACIDS_AND_BILE_SALTS_VIA_24_HYDROXYCHOLESTEROL
CNS_regionOptic_pathway
REACTOME_SIGNALING_BY_MAPK_MUTANTS
CNS_regionOther
REACTOME_DEPOSITION_OF_NEW_CENPA_CONTAINING_NUCLEOSOMES_AT_THE_CENTROMERE
REACTOME_OXIDATIVE_STRESS_INDUCED_SENESCENCE
molecular_subtypeLGG_NF1_germline
raceNot_Hispanic_or_Latino
REACTOME_DISSOLUTION_OF_FIBRIN_CLOT
age_at_diagnosis_days
molecular_subtypeLGG_RTK
raceNot_Available
REACTOME_KERATAN_SULFATE_DEGRADATION
REACTOME_REGULATION_OF_TLR_BY_ENDOGENOUS_LIGAND
REACTOME_CASPASE_ACTIVATION_VIA_DEATH_RECEPTORS_IN_THE_PRESENCE_OF_LIGAND
CNS_regionPosterior_fossa
CNS_regionMixed
REACTOME_REGULATION_OF_COMMISSURAL_AXON_PATHFINDING_BY_SLIT_AND_ROBO
REACTOME_SCAVENGING_BY_CLASS_A_RECEPTORS
−4
−2
0
2

FIG 5. (A-C) Plots of generalized linear model (GLM) coefficients indicating the top pathways selected in the elastic net classification approach for prediction of each imaging cluster vs the rest of the tumor cohort.

Survival Differences Across Imaging Clusters

Cox regression analysis (FIG 6A), adjusted for tumor resection extent and interactions between molecular subgroups and imaging cluster covariates, revealed significantly better EFS for cluster 3 compared to cluster 1 (reference group). However, Kaplan-Meier survival analysis and log-rank tests showed no significant survival differences across clusters ($p > 0.05$).

BRAF-fusion status influenced survival outcomes in cluster 3 (Supplementary FIGs 5-8). Within cluster 3, tumors with *KIAA1549::BRAF* fusions had significantly worse EFS compared to other cluster 3 tumors (HR = 47, $p = 0.012$; FIG 6A). For patients without *BRAF*-fusion tumors, cluster 3 had significantly better EFS than cluster 1 (HR = 0.22, $p = 0.002$; Supplementary FIG 6). However, no survival benefit was observed in cluster 3 relative to cluster 1 among patients with *BRAF*-fusion tumors (HR = 1.3, $p = 0.573$; Supplementary FIG 8).

We further analyzed *KIAA1549::BRAF* fusion breakpoints, categorizing them as common (exons 15:09, 16:09, 16:11, and 18:10) or rare (all others) following prior work[27]. Breakpoint types were significantly associated with imaging clusters (Fisher exact $p = 2.0e\text{-}03$), with rare breakpoints exclusively in cluster 3 (FIG 6B, Supplementary FIG 9). Patients with rare breakpoints had worse EFS compared to those with common breakpoints in a log-rank test ($p = 0.03$; FIG 6C), but the difference was not significant in a Cox regression model accounting for resection extent and imaging cluster (HR = 1.7, $p = 0.48$; Supplementary FIG 10).

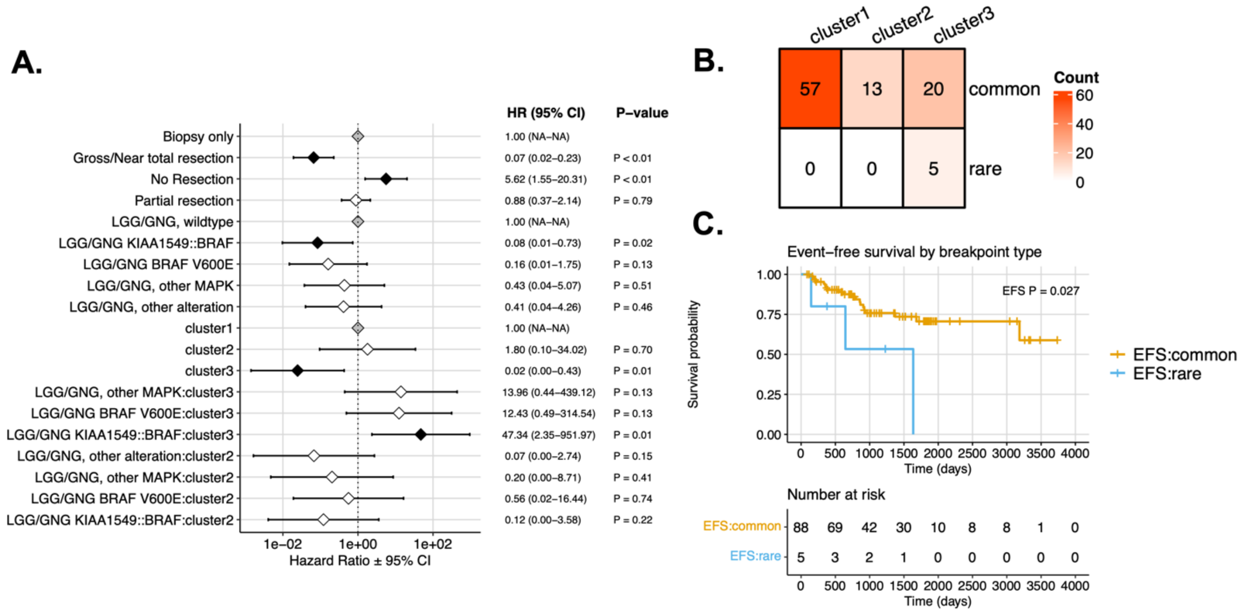


FIG 6. Event-free survival (EFS) by imaging cluster. (A) Cox regression model forest plot of EFS, including covariates for extent of tumor resection, molecular subtype, imaging cluster, and subtype-imaging cluster interaction terms. Gray points indicate reference levels (Biopsy only, LGG/GNG wildtype, and imaging Cluster 1), and black points indicate terms with statistically significant hazard ratios relative to reference levels. (B) Distribution of *KIAA1549::BRAF* fusion tumor breakpoint types and corresponding enrichment odds ratios. (C) Kaplan-Meier plot of EFS among patients with *KIAA1549::BRAF* fusion tumors by *KIAA1549::BRAF* breakpoint type.

DISCUSSION

In this manuscript, we presented an imaging-based clustering method for pLGGs, identifying three subgroups with different molecular characteristics, extending beyond the classic WHO CNS classifications. These clusters revealed distinct transcriptomic profiles, offering insights into the molecular underpinnings of glioma and potential therapeutic targets. Cluster 1 showed mechanisms driving tumor growth[35, 39, 40], including *PDGFRB* expression, induced by microglia in low-grade glioma which enhances migratory capacity[5, 32]. *KIAA1549::BRAF* pilocytic astrocytomas, enriched in cluster 1, have been found to aberrantly activate oxidative phosphorylation in part through deleterious mitochondrial DNA mutations [41, 42]. Elevated IL10 levels, typically found in invasive higher grade gliomas that suppress anti-tumor immune activity, highlight the potential for targeted therapies using direct inhibitors or neutralizing antibodies [35].

Cluster 2 showed features of metabolic plasticity and glioma stem cell maintenance, common in higher-grade gliomas [31], highlighting therapeutic potential of inhibiting glycolytic and fatty acid oxidation pathways [29, 30, 43]. Differential expression in adhesion and extracellular matrix remodeling pathways between clusters 2 and 3 suggests variations in microenvironmental remodeling, potentially influencing tumor spread, invasiveness, immune escape, and response to immune checkpoint blockade [44]. Cluster 3 showed upregulation of pathways related to histone acetylation and DNA methylation, potentially linked to *BRAF V600E* mutation in low-grade gliomas [45]. Epigenetic reprogramming, observed in *BRAF V600E*-driven melanomas and colorectal cancers, enables resistance to *BRAF* inhibitors overcomes *BRAF* dependency, counters oncogene-induced senescence, highlighting the need for combination therapies targeting BET proteins or histone deacetylases alongside the oncogenic *BRAF* mutation [45-47].

Rare/novel *KIAA1549::BRAF* fusion breakpoints (*KIAA1549::BRAF* 15:11) were only found in cluster 3, indicating similar imaging characteristics among these patients. Consistent with previous studies [6, 27], tumors with rare/novel *KIAA1549::BRAF* fusion breakpoints, highly enriched in recurrent/progressive pLGGs, exhibit a poorer prognosis. Despite being associated with poorer prognoses, these cases were categorized under cluster 3 with overall favorable outcomes, suggesting the need for larger studies to refine risk stratification. Notably, the hazard ratio for *KIAA1549::BRAF* tumors in cluster 3 was higher than that of *BRAF V600E* tumors, which is not in agreement with the established prognosis of these tumors [48]. This discrepancy may be due to the inclusion of rare/novel *KIAA1549::BRAF* fusion breakpoints in cluster 3, and the fact that 8 out of 12 tumors with the *BRAF V600E* mutation in this group had not progressed during the time period of this study.

Radiomic analyses in pLGGs have primarily focused on predicting single genetic alterations, such as the *BRAF V600E* mutation or *KIAA1549::BRAF* fusion [13-16], limiting clinical utility as molecular data is being increasingly incorporated in diagnostics per 2021 WHO CNS 5th Edition recommendations [4]. Unlike other studies on pLGGs, our work explored broader associations between imaging phenotypes and underlying genomic or transcriptomic pathways, providing a comprehensive radiogenomic framework. While similar clustering approaches have been applied in neurodegenerative diseases and adult glioblastoma [49, 50], our study is the first to extend this method to pediatric neuro-oncology.

Our study faced challenges due to limited availability of imaging and genomic data, compounded by the rarity of pediatric brain tumors and a small cohort size, particularly for non-pilocytic astrocytomas in specific brain sites. Survival analysis relied on subjective progression assessments from clinical notes and MRI reviews. Variability in tumor location and molecular subtypes between the training and test sets may have impacted the performance of supervised models. Furthermore, molecular data obtained from localized tumor samples may not fully reflect the tumor's broader molecular heterogeneity, limiting its correlation with radiomic features. Additionally, other critical data layers, such as methylation profiling, metabolomics, proteomics, and pathomics, were unavailable, limiting deeper analyses. Future efforts should focus on larger multi-institutional cohorts to test generalizability, integration of multi-modal data to enhance tumor subtype differentiation, and identification of therapeutic targets.

CONCLUSIONS

In conclusion, our unsupervised clustering of treatment-naïve imaging and clinical highlights potential for personalized pLGG treatments beyond WHO-recognized molecular subtypes. This approach supports molecular tumor boards on pLGGs in developing more targeted treatment strategies, moving beyond conventional methods.

REFERENCES

1. 1. Kim Y-A, Cho D-Y, Przytycka TM. Understanding genotype-phenotype effects in cancer via network approaches. *PLoS computational biology* 2016;12:e1004747
2. 2. Gatenby RA, Grove O, Gillies RJ. Quantitative imaging in cancer evolution and ecology. *Radiology* 2013;269:8-14
3. 3. Ostrom QT, Price M, Ryan K, et al. CBTRUS statistical report: pediatric brain tumor foundation childhood and adolescent primary brain and other central nervous system tumors diagnosed in the United States in 2014–2018. *Neuro-oncology* 2022;24:iii1-iii38
4. 4. Louis DN, Perry A, Wesseling P, et al. The 2021 WHO classification of tumors of the central nervous system: a summary. *Neuro-oncology* 2021;23:1231-1251
5. 5. Ryall S, Tabori U, Hawkins C. Pediatric low-grade glioma in the era of molecular diagnostics. *Acta neuropathologica communications* 2020;8:1-22
6. 6. Ryall S, Zapotocky M, Fukuoka K, et al. Integrated molecular and clinical analysis of 1,000 pediatric low-grade gliomas. *Cancer cell* 2020;37:569-583. e565
7. 7. Krishnatry R, Zhukova N, Guerreiro Stucklin AS, et al. Clinical and treatment factors determining long-term outcomes for adult survivors of childhood low-grade glioma: a population-based study. *Cancer* 2016;122:1261-1269
8. 8. Manoharan N, Liu KX, Mueller S, et al. Pediatric low-grade glioma: Targeted therapeutics and clinical trials in the molecular era.

Neoplasia 2023;36:100857

9. 9. Jones DT, Kieran MW, Bouffet E, et al. Pediatric low-grade gliomas: next biologically driven steps. *Neuro-oncology* 2018;20:160-173

10. 10. Fangusaro J, Jones DT, Packer RJ, et al. Pediatric low-grade glioma: State-of-the-art and ongoing challenges. *Neuro-oncology* 2024;26:25-37

11. 11. Madhogarhia R, Haldar D, Bagheri S, et al. Radiomics and radiogenomics in pediatric neuro-oncology: A review. *Neuro-Oncology Advances* 2022;4:vdac083

12. 12. Fathi Kazerooni A, Bakas S, Saligheh Rad H, et al. Imaging signatures of glioblastoma molecular characteristics: a radiogenomics review. *Journal of Magnetic Resonance Imaging* 2020;52:54-69

13. 13. Trasolini A, Erker C, Cheng S, et al. MR Imaging of Pediatric Low-Grade Gliomas: Pretherapeutic Differentiation of BRAF V600E Mutation, BRAF Fusion, and Wild-Type Tumors in Patients without Neurofibromatosis-1. *American Journal of Neuroradiology* 2022;43:1196-1201

14. 14. Tak D, Ye Z, Zapaishchykova A, et al. Noninvasive molecular subtyping of pediatric low-grade glioma with self-supervised transfer learning. *medRxiv* 2023

15. 15. Kudus K, Wagner MW, Namdar K, et al. Increased confidence of radiomics facilitating pretherapeutic differentiation of BRAF-altered pediatric low-grade glioma. *European Radiology* 2023:1-10

16. 16. Wagner M, Hainc N, Khalvati F, et al. Radiomics of pediatric low-grade gliomas: toward a pretherapeutic differentiation of BRAF-mutated and BRAF-fused tumors. *American Journal of Neuroradiology* 2021;42:759-765

17. 17. Haldar D, Kazerooni AF, Arif S, et al. Unsupervised machine learning using K-means identifies radiomic subgroups of pediatric low-grade gliomas that correlate with key molecular markers. *Neoplasia* 2023;36:100869

18. 18. Wen J, Varol E, Yang Z, et al. Subtyping Brain Diseases from Imaging Data. In: Colliot O, ed. *Machine Learning for Brain Disorders*. New York, NY: Springer US; 2023:491-510

19. 19. Lilly JV, Rokita JL, Mason JL, et al. The children's brain tumor network (CBTN)-Accelerating research in pediatric central nervous system tumors through collaboration and open science. *Neoplasia* 2023;35:100846

20. 20. Shapiro JA, Gaonkar KS, Spielman SJ, et al. OpenPBTA: The Open Pediatric Brain Tumor Atlas. *Cell Genomics* 2023

21. 21. Fathi Kazerooni A, Arif S, Madhogarhia R, et al. Automated tumor segmentation and brain tissue extraction from multiparametric MRI of pediatric brain tumors: A multi-institutional study. *Neuro-Oncology Advances* 2023;5:vdad027

22. 22. Vossough A, Khalili N, Familiar AM, et al. Training and Comparison of nnU-Net and DeepMedic Methods for Autosegmentation of Pediatric Brain Tumors. *American Journal of Neuroradiology* 2024

23. 23. Prasanna P, Tiwari P, Madabhushi A. Co-occurrence of local anisotropic gradient orientations (CoLlAGe): a new radiomics descriptor. *Scientific reports* 2016;6:37241

24. 24. Durinck S, Spellman PT, Birney E, et al. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nature protocols* 2009;4:1184-1191

25. 25. Durinck S, Moreau Y, Kasprzyk A, et al. BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. *Bioinformatics* 2005;21:3439-3440

26. 26. Lawrence M, Huber W, Pagès H, et al. Software for computing and annotating genomic ranges. *PLoS computational biology* 2013;9:e1003118

27. 27. Haas-Kogan DA, Aboian MS, Minturn JE, et al. Everolimus for children with recurrent or progressive low-grade glioma: Results from the phase II PNOC001 trial. *Journal of Clinical Oncology* 2024;42:441-451

28. 28. Project tSJCsRHWUPCG. Whole-genome sequencing identifies genetic alterations in pediatric low-grade gliomas. *Nature genetics* 2013;45:602-612

29. 29. Lin H, Patel S, Affleck VS, et al. Fatty acid oxidation is required for the respiration and proliferation of malignant glioma cells. *Neuro-oncology* 2017;19:43-54

30. 30. McKelvey KJ, Wilson EB, Short S, et al. Glycolysis and fatty acid oxidation inhibition improves survival in glioblastoma. *Frontiers in Oncology* 2021;11:633210

31. 31. Byrne KF, Pal A, Curtin JF, et al. G-protein-coupled receptors as therapeutic targets for glioblastoma. *Drug Discovery Today* 2021;26:2858-2870

32. 32. Wallmann T, Zhang X-M, Wallerius M, et al. Microglia induce PDGFRB expression in glioma cells to enhance their migratory capacity. *IScience* 2018;9:71-83

33. 33. Lucas C-HG, Gupta R, Doo P, et al. Comprehensive analysis of diverse low-grade neuroepithelial tumors with FGFR1 alterations reveals a distinct molecular signature of rosette-forming glioneuronal tumor. *Acta neuropathologica communications* 2020;8:1-17

34. 34. Becker AP, Scapulatempo-Neto C, Carloni AC, et al. KIAA1549: BRAF gene fusion and FGFR1 hotspot mutations are prognostic factors in pilocytic astrocytomas. *Journal of Neuropathology & Experimental Neurology* 2015;74:743-754

35. 35. Widodo SS, Dinevska M, Furst LM, et al. IL-10 in glioma. *British Journal of Cancer* 2021;125:1466-1476

36. 36. Liang R, Wu C, Liu S, et al. Targeting interleukin-13 receptor α2 (IL-13Rα2) for glioblastoma therapy with surface functionalized nanocarriers. *Drug Delivery* 2022;29:1620-1630

37. 37. Joshi BH, Leland P, Lababidi S, et al. Interleukin-4 receptor alpha overexpression in human bladder cancer correlates with the pathological grade and stage of the disease. *Cancer medicine* 2014;3:1615-1628

38. 38. Todaro M, Lombardo Y, Francipane M, et al. Apoptosis resistance in epithelial tumors is mediated by tumor-cell-derived interleukin-4. *Cell Death & Differentiation* 2008;15:762-772

39. 39. Bhardwaj R, Suzuki A, Leland P, et al. Identification of a novel role of IL-13Rα2 in human Glioblastoma multiforme: interleukin-13 mediates signal transduction through AP-1 pathway. *Journal of Translational Medicine* 2018;16:1-13

40. 40. Mintz A, Gibo DM, Slagle-Webb B, et al. IL-13Rα2 is a glioma-restricted receptor for interleukin-13. *Neoplasia* 2002;4:388-399
41. 41. Picard D, Felsberg J, Langini M, et al. Integrative multi-omics reveals two biologically distinct groups of pilocytic astrocytoma. *Acta Neuropathologica* 2023:1-14
42. 42. Lueth M, Wronski L, Giese A, et al. Somatic mitochondrial mutations in pilocytic astrocytoma. *Cancer genetics and cytogenetics* 2009;192:30-35
43. 43. Kant S, Kesarwani P, Prabhu A, et al. Enhanced fatty acid oxidation provides glioblastoma cells metabolic plasticity to accommodate to its dynamic nutrient microenvironment. *Cell Death & Disease* 2020;11:253
44. 44. Wang L-j, Lv P, Lou Y. A Novel TAF-Related Signature Based on ECM Remodeling Genes Predicts Glioma Prognosis. *Frontiers in Oncology* 2022;12:862723
45. 45. Khaliq M, Fallahi-Sichani M. Epigenetic mechanisms of escape from BRAF oncogene dependency. *Cancers* 2019;11:1480
46. 46. Yu Y, Schleich K, Yue B, et al. Targeting the senescence-overriding cooperative activity of structurally unrelated H3K9 demethylases in melanoma. *Cancer cell* 2018;33:322-336. e328
47. 47. Roesch A, Fukunaga-Kalabis M, Schmidt EC, et al. A temporarily distinct subpopulation of slow-cycling melanoma cells is required for continuous tumor growth. *Cell* 2010;141:583-594
48. 48. Lassaletta A, Zapotocky M, Mistry M, et al. Therapeutic and prognostic implications of BRAF V600E in pediatric low-grade gliomas. *Journal of Clinical Oncology* 2017;35:2934-2941
49. 49. Rathore S, Akbari H, Rozycki M, et al. Radiomic MRI signature reveals three distinct subtypes of glioblastoma with different clinical and molecular characteristics, offering prognostic value beyond IDH1. *Scientific reports* 2018;8:5087
50. 50. Guo J, Fathi Kazerooni A, Toorens E, et al. Integrating imaging and genomic data for the discovery of distinct glioblastoma subtypes: a joint learning approach. *Scientific Reports* 2024;14:4922

SUPPLEMENTAL FILES

SI1. Image Preprocessing and Tumor Subregion Segmentation

In this research, we standardized images prior to extracting radiomic features. Initially, MRI scans from each patient were oriented to the left-posterior-superior (LPS) coordinate system. The T1-weighted (T1w), T2-weighted (T2w), and T2-FLAIR, were aligned with the corresponding post-contrast T1w (T1w-Gd) images. These images were then uniformly resampled to a 1 mm^3 isotropic resolution using the anatomical SRI24 atlas through the Greedy algorithm, as facilitated by the open-source Cancer Imaging Phenomics Toolkit (CaPTk, version 1.8.1, available at https://www.cbica.upenn.edu/captk) [1,2]. Following this, the images underwent skull stripping using a specialized pediatric brain tissue extraction tool [3]. After removing outlier pixels that exceeded the 99.9th percentile in the image histogram, the skull-stripped images were normalized to an intensity range of [0, 255]. Brain tumor segmentation was conducted with our proprietary pediatric brain tumor segmentation tool [4], complemented by manual adjustments where needed. This tool delineates various tumor subregions, such as the enhancing tumor, non-enhancing tumor, cyst, and edema, which are collectively used to create a complete lesion segmentation, as illustrated in FIG 1.

SI2. Enrichment Analysis of Molecular Pathways

Differentially expressed genes were used as input for gene set enrichment analysis (GSEA [5]), and differentially regulated pathways were defined as having a Benjamini-Hochberg corrected p-value of less than 0.05. Pathway annotations were derived from Reactome [6] using the Molecular Signatures Database (MSigDb [7]) v2023.1.Hs 74. The ClusterProfiler [8] R package was used to generate visualizations of differentially regulated pathways. To understand differences in gene co-expression networks, we investigated cohort-wide gene-gene correlations subject to a scale-free topology using CEMITool [9], which internally leverages GSEA to determine differences in network module expression as a function of a categorical annotation (in this case, imaging cluster). We also directly investigated differentially co-expressed pathways using Gene Set Net Correlation Analysis (GSNCA [10]), which estimates gene-specific weights based on a gene's cross-correlation with all other genes in a given pathway. We focused on the identification of differentially co-expressed pathways and hub genes, defined as a gene with the greatest weight in a given pathway.

To identify differentially expressed biological pathways across imaging groups, we filtered gene expression data for the 'protein coding' gene type. We then provided expected counts as input to DeSeq2 [11] for differential gene expression analysis, specifying imaging cluster as the factor variable in the model's design. Lists of differentially expressed genes derived from pairwise comparisons across all three imaging clusters were ranked and used as input for pre-ranked GSEA [12] to identify differentially expressed biological pathways. Pathway annotations were derived from Reactome using the Molecular Signatures Database v2023.1.Hs [7].

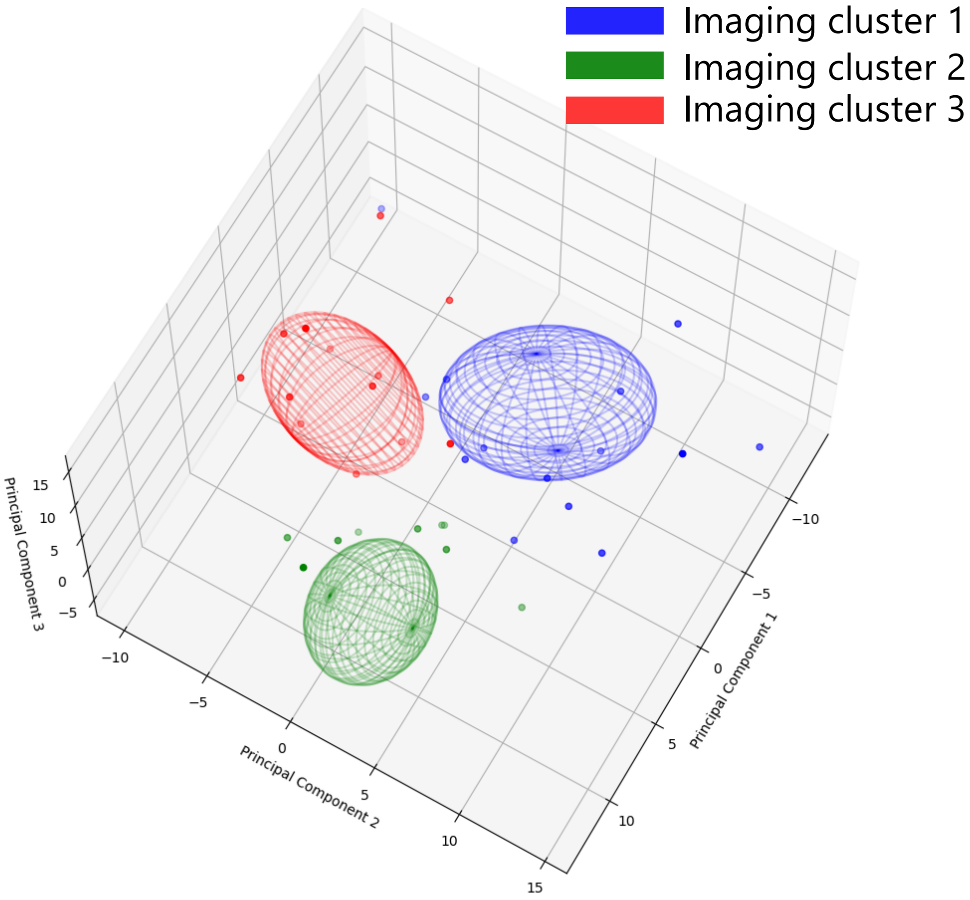
To characterize and profile gene co-expression sub-networks in low grade gliomas, we used the CEMITool bioconductor package [9] on TPM gene expression data. CEMITool defines co-expressed modules based on the correlation structure and a re-scaled gene-gene adjacency metric to define subnetworks following a scale-free topology. GSEA and over-representation analyses are then used to characterize cluster-specific subnetwork differences.

Supplementary Table 1. Summary of demographics and clinical characteristics of the pLGG patients across the three imaging clusters (collected through the CBTN data repository).

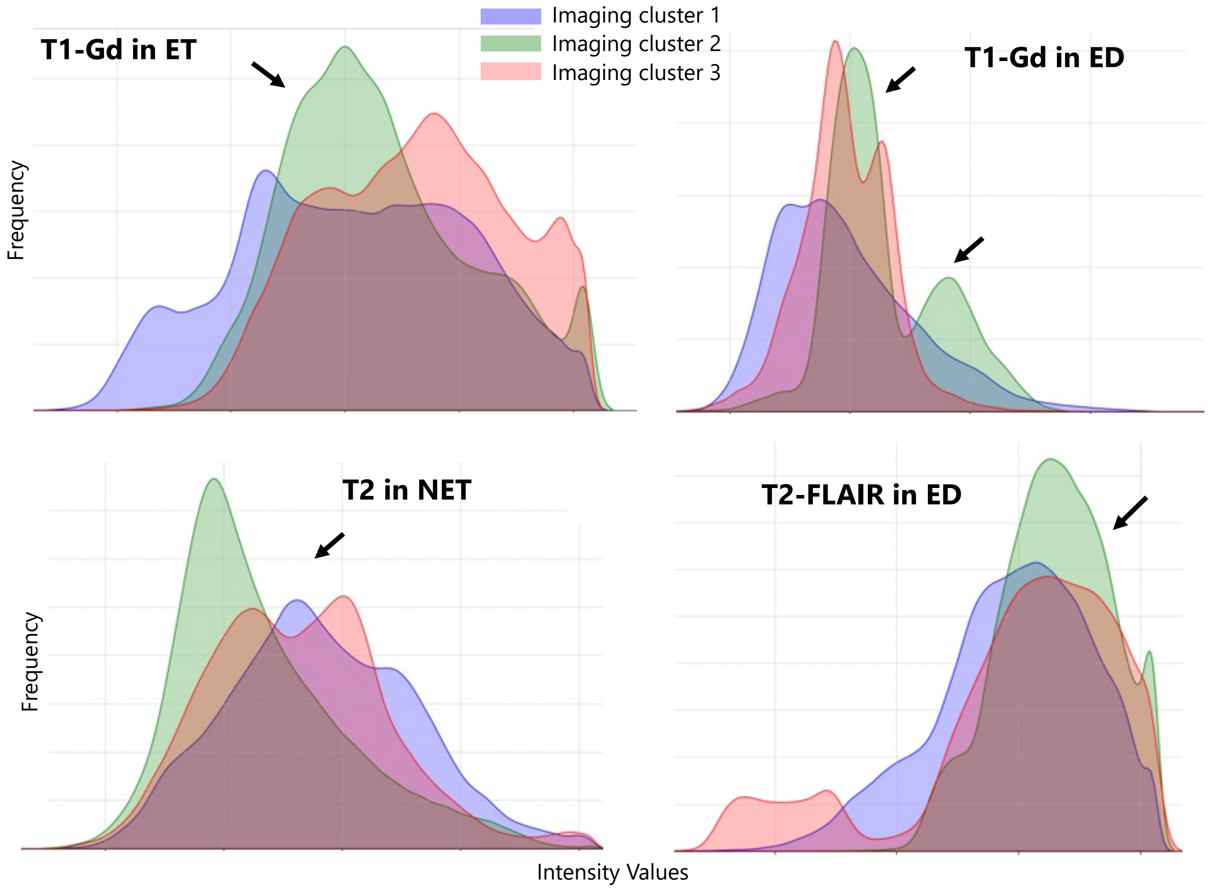
Variable	Value	# of Patients			
		Cluster 1	Cluster 2	Cluster 3	P-Values
Cohort Size		78	52	71	
Age Range (months)		7.5 – 238.3; Mean, 95.1	6.7 – 280.7; Mean, 115.4	4.3 – 279.5; Mean, 117.0	0.070
Sex					0.677
	Female	35	25	37	
	Male	43	27	34	
Tumor Location					3.096E-11
	Basal Ganglia	1	1	0	
	Brainstem	1	5	4	
	Cerebellar	54	16	16	
	Intraventricular	1	3	1	
	Lobar (frontal, parietal, temporal, or occipital lobes)	8	17	19	
	Multi-regional	3	8	3	
	Suprasellar	5	2	27	
	Thalamus	5	0	1	
Extent of Tumor Resection					0.0004
	Gross or near total resection	53	27	29	
	Partial resection	22	9	27	
	Biopsy	3	12	12	
	Not Reported/Unavailable	0	4	3	
Treatment					0.035
	Radiotherapy	1	3	2	
	Chemotherapy	10	8	24	
	Radiotherapy and Chemotherapy	2	1	1	
	None or Not Available	65	40	44	
WHO 2021 Classification					0.005
	Diffuse astrocytoma, MYB- or MYBL1-altered	1	2	0	
	Diffuse glioneuronal tumor with oligodendroglioma-like features and nuclear clusters	0	0	1	
	Diffuse low-grade glioma, MAPK pathway-altered	57	25	46	
	Ganglioglioma	1	3	1	
	Glioneuronal and neuronal tumors	0	3	0	
	Pediatric-type diffuse low-grade gliomas, NOS	6	13	14	
	Pilocytic astrocytoma	13	4	8	
	Subependymal giant cell astrocytoma	0	2	1	

Event-Free Survival (EFS) Range (months)		3.83 – 127.1; Mean, 39.5	3.0 – 124.8; Mean, 42.1	2.83 – 133.4; Mean, 38.7	0.853
Patients with EFS Outcome		21	15	25	N/A

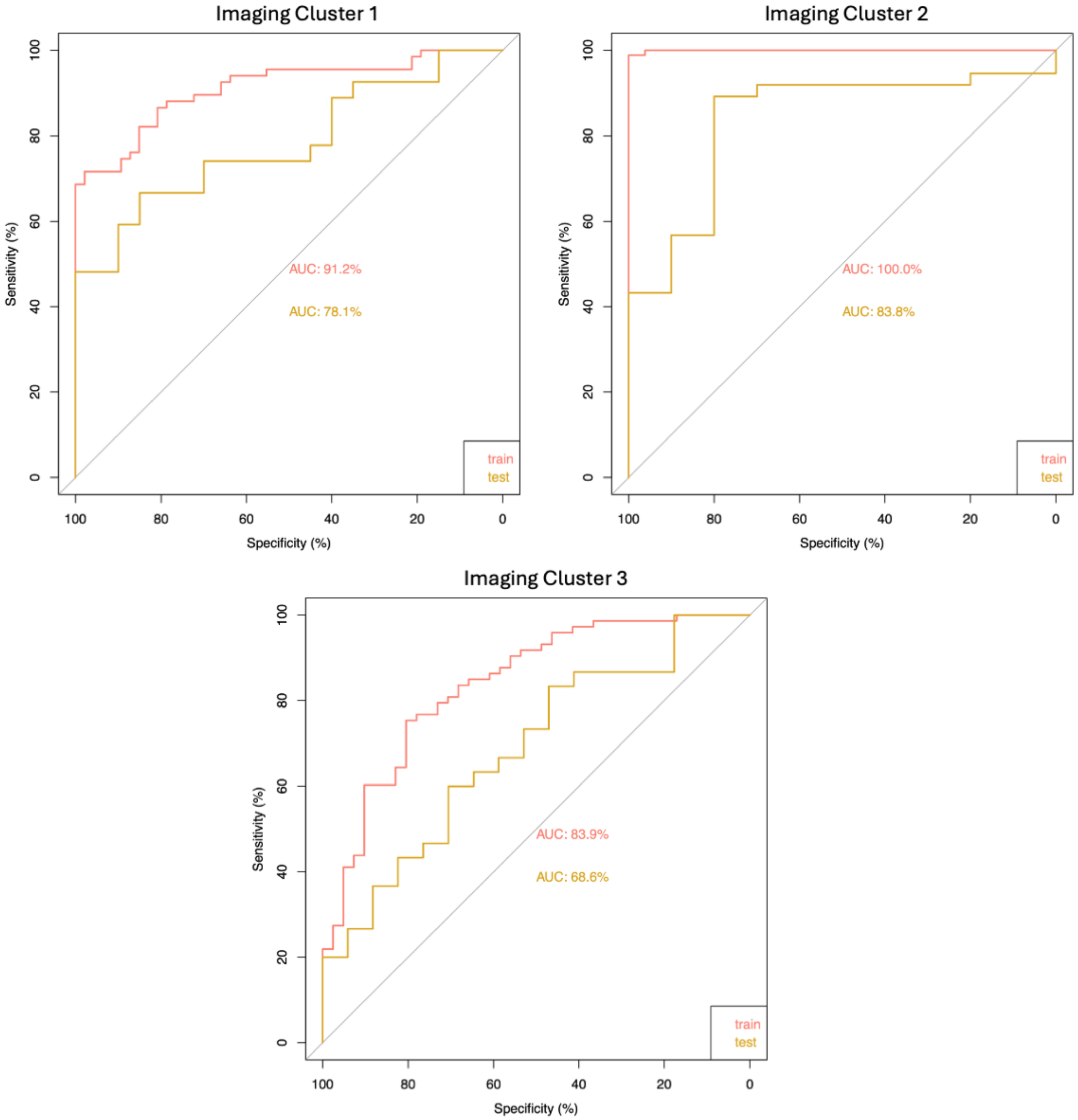
Note: For categorical variables, including Sex, Tumor Location, Extent of Tumor Resection, Treatment, and WHO 2021 Classification, Chi-squared test was used to calculate the differences across the imaging clusters. For continuous variables, including Age and Progression-Free Survival, Kruskal-Wallis test was used to compare the differences across the imaging clusters.



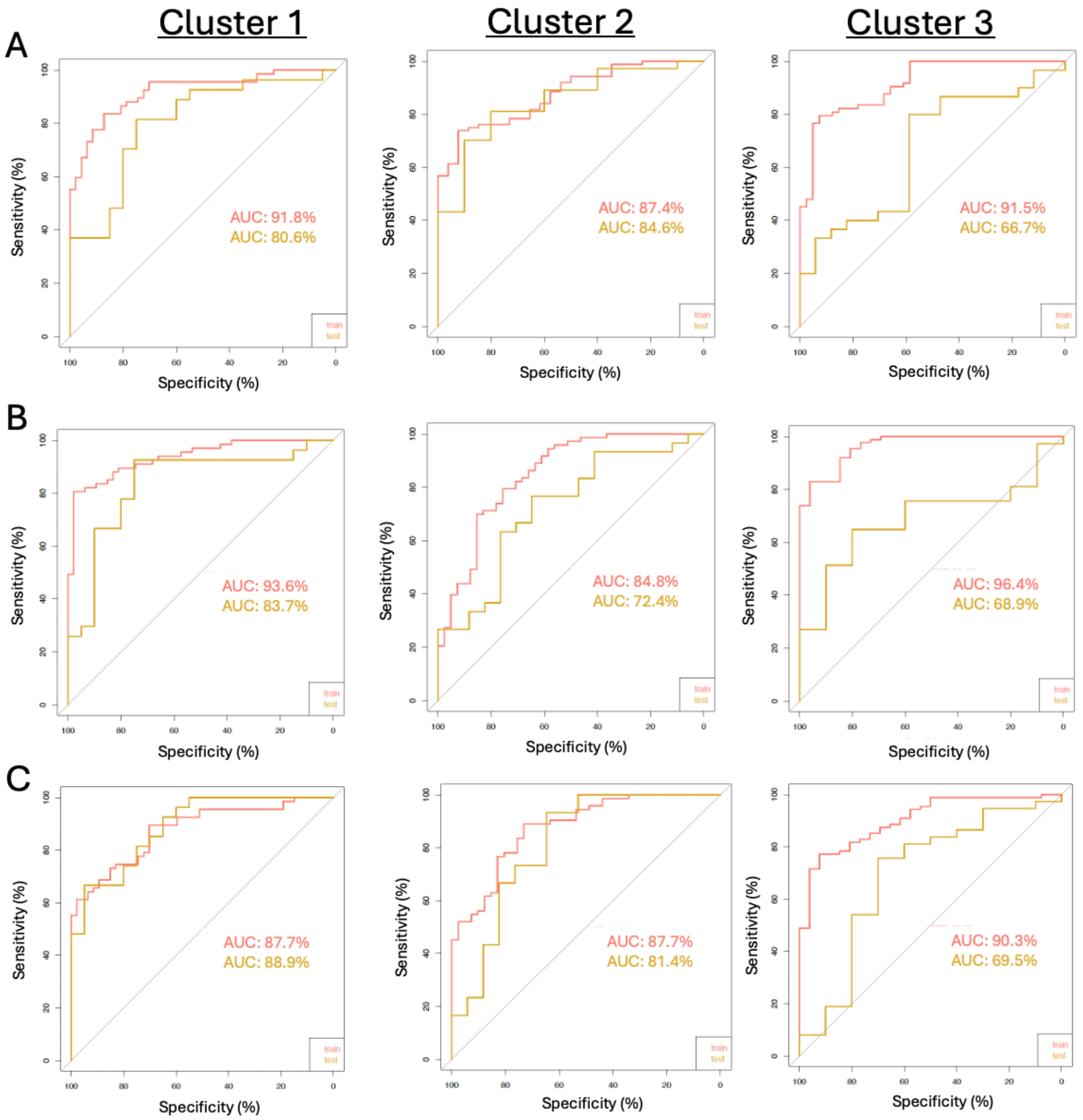
Supplementary FIG 1. Three-dimensional visualization of imaging clusters in the replication set.



Supplementary FIG 2. Histograms of image intensities within different tumorous subregions, i.e., enhancing tumor (ET), edema (ED), non-enhancing tumor (NET), for the three imaging clusters.



Supplementary FIG 3. ROC curves illustrating the performances of Elastic-Net regression models applied to transcriptomic pathways to classify each imaging cluster from the rest. This analysis compares the relative association of differentially expressed/co-expressed signaling pathways, molecular subtypes, and clinical variables with imaging clusters.

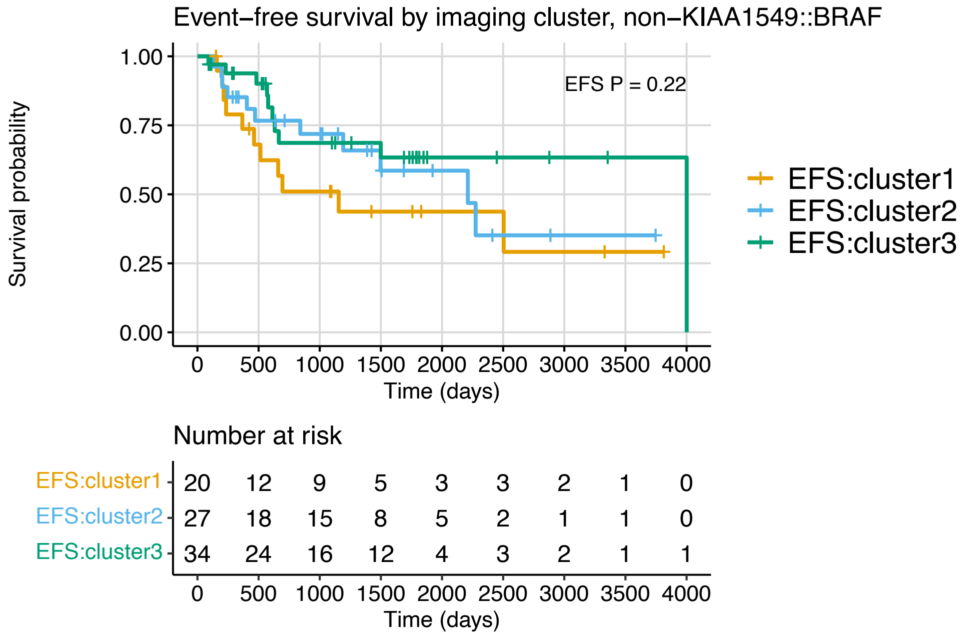


Supplementary FIG 4. ROC curves indicating the performances of Elastic-Net regression models applied to transcriptomic pathways for three additional independent training-testing splits of the data.

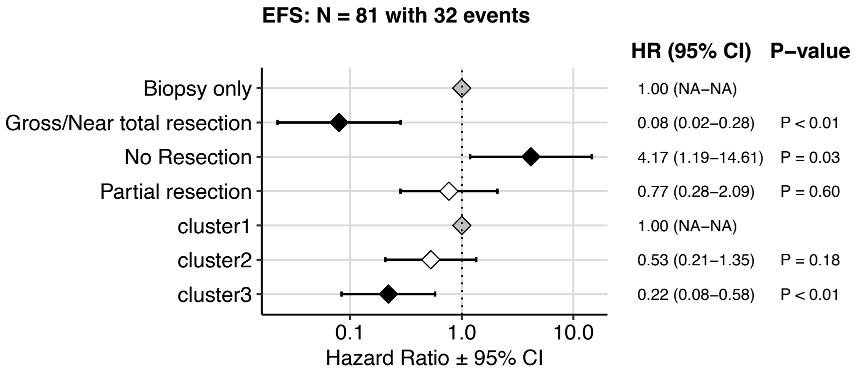
Supplementary Table 2. Summary of Poisson generalized linear model parameters comparing counts across imaging clusters and molecular subtypes.

Characteristic	IRR[1]	95% CI[1]	p-value
2021_WHO_Classification			
Pediatric-type diffuse low-grade gliomas, NOS	-	-	
Diffuse low-grade glioma, MAPK pathway-altered	0.64	0.49, 0.83	<0.001
Ganglioglioma	0.07	0.00, 0.32	0.008
Pilocytic astrocytoma	3.00	2.60, 3.49	<0.001
Imaging_Cluster			
1	-	-	
2	0.36	0.23, 0.53	<0.001
3	0.86	0.69, 1.06	0.2
2021_WHO_Classification * Imaging_Cluster			
Diffuse low-grade glioma, MAPK pathway-altered * 2	4.67	2.91, 7.75	<0.001
Ganglioglioma * 2	11.2	2.15, 207	0.022
Pilocytic astrocytoma * 2	0.6	0.38, 0.98	0.034
Diffuse low-grade glioma, MAPK pathway-altered * 3	2.85	2.08, 3.94	<0.001
Ganglioglioma * 3	3.5	0.65, 64.9	0.2
Pilocytic astrocytoma * 3	0.5	0.39, 0.64	<0.001

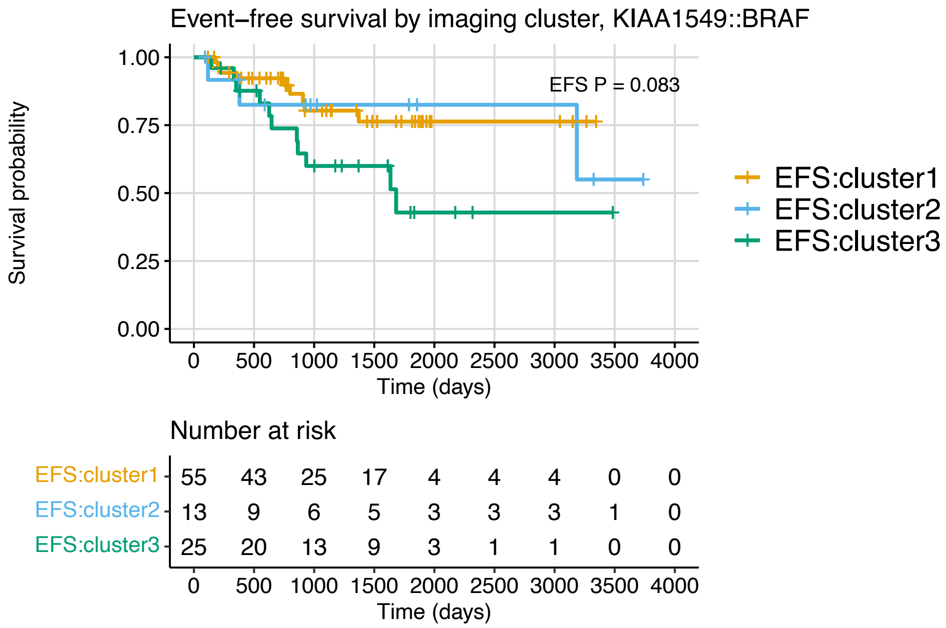
[1] IRR = Incidence Rate Ratio, CI = Confidence Interval



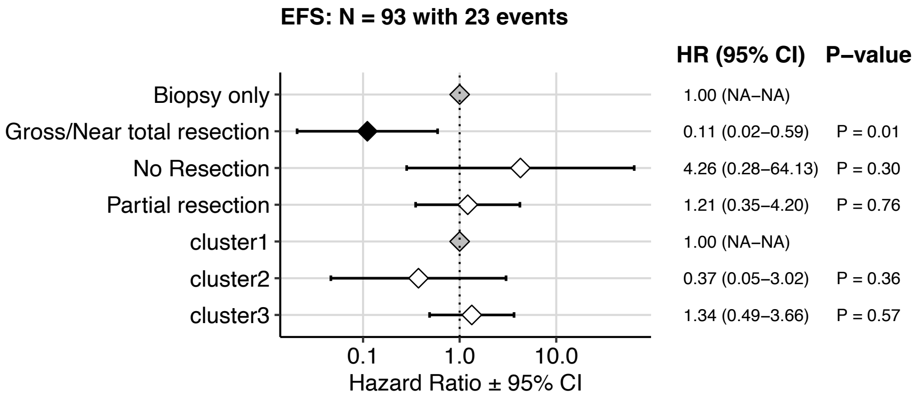
Supplementary FIG 5. Kaplan Meier plots of event-free survival by imaging cluster in non-KIAA1549::BRAF fusion tumors.



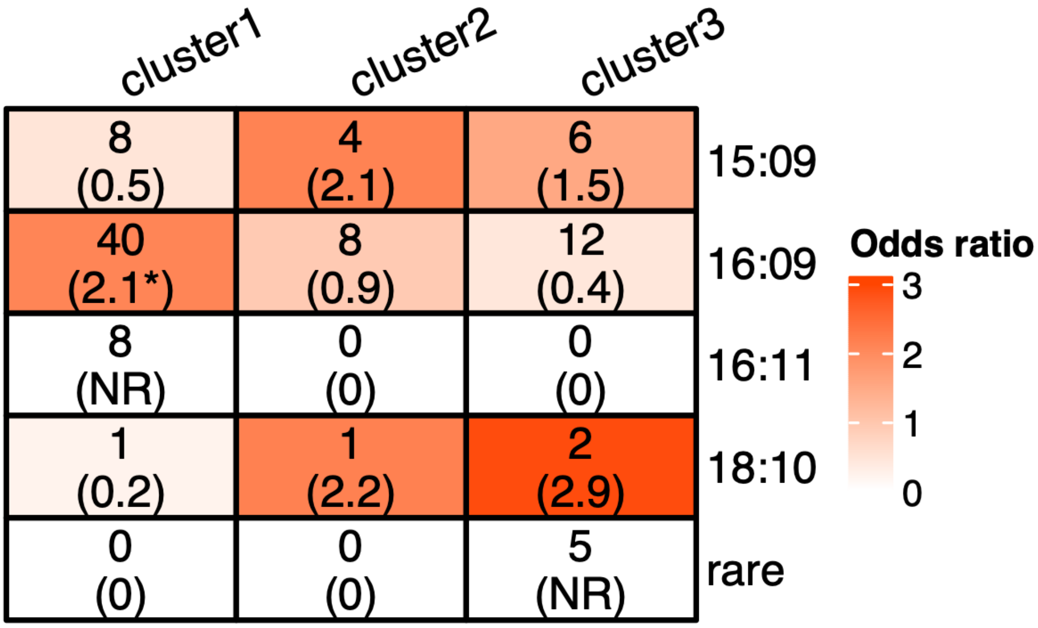
Supplementary FIG 6. Cox regression model forest plots of event-free survival by imaging cluster in non-KIAA1549::BRAF fusion tumors including covariates for extent of tumor resection and imaging cluster.



Supplementary FIG 7. Kaplan Meier plots of event-free survival by imaging cluster in KIAA1549::BRAF fusion tumors.

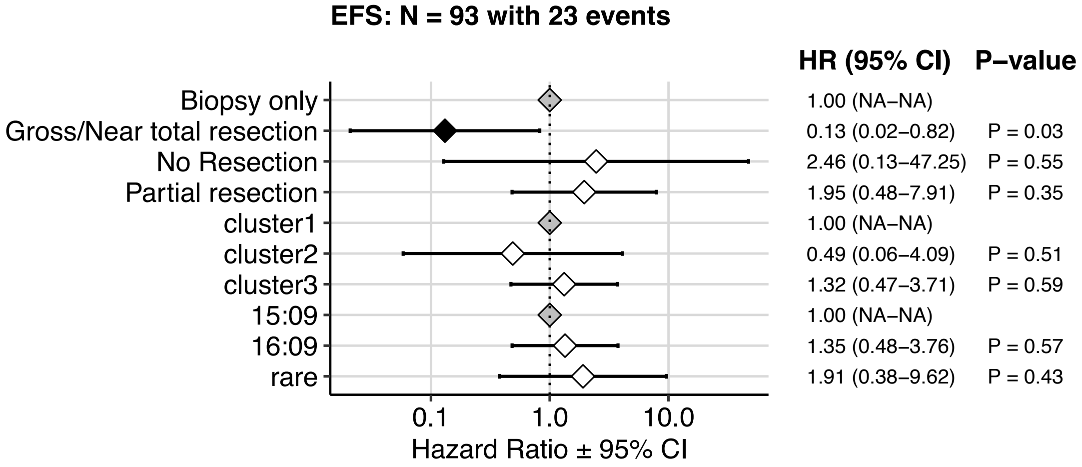


Supplementary FIG 8. Cox regression model forest plots of event-free survival by imaging cluster in KIAA1549::BRAF fusion tumors including covariates for extent of tumor resection and imaging cluster.



	cluster1	cluster2	cluster3
15:09	8 (0.5)	4 (2.1)	6 (1.5)
16:09	40 (2.1*)	8 (0.9)	12 (0.4)
16:11	8 (NR)	0 (0)	0 (0)
18:10	1 (0.2)	1 (2.2)	2 (2.9)
rare	0 (0)	0 (0)	5 (NR)

Supplementary FIG 9. Distribution of KIAA1549::BRAF fusion tumor breakpoint groups among imaging clusters, and corresponding enrichment odds ratios in parentheses. *$p<0.05$. NR = Not reportable.



	HR (95% CI)	P–value
Biopsy only	1.00 (NA–NA)	
Gross/Near total resection	0.13 (0.02–0.82)	P = 0.03
No Resection	2.46 (0.13–47.25)	P = 0.55
Partial resection	1.95 (0.48–7.91)	P = 0.35
cluster1	1.00 (NA–NA)	
cluster2	0.49 (0.06–4.09)	P = 0.51
cluster3	1.32 (0.47–3.71)	P = 0.59
15:09	1.00 (NA–NA)	
16:09	1.35 (0.48–3.76)	P = 0.57
rare	1.91 (0.38–9.62)	P = 0.43

Supplementary FIG 10. Cox regression model forest plot of EFS among patients with KIAA1549::BRAF fusion tumors, including covariates for extent of tumor resection, imaging cluster, and breakpoint group. Gray points indicate reference levels for each covariate, and black points indicate terms with statistically significant hazard ratios relative to reference levels.

Supplementary References

51. **1.** Davatzikos C, Rathore S, Bakas S, et al. Cancer imaging phenomics toolkit: quantitative imaging analytics for precision diagnostics and predictive modeling of clinical outcome. *Journal of medical imaging.* 2018; 5(1):011018-011018.

52. **2.** Pati S, Singh A, Rathore S, et al. The cancer imaging phenomics toolkit (CaPTk): technical overview. Paper presented at: Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 5th International Workshop, BrainLes 2019, Held in Conjunction with MICCAI 2019, Shenzhen, China, October 17, 2019, Revised Selected Papers, Part II 52020.

53. **3.** Fathi Kazerooni A, Arif S, Madhogarhia R, et al. Automated tumor segmentation and brain tissue extraction from multiparametric MRI of pediatric brain tumors: A multi-institutional study. *Neuro-Oncology Advances.* 2023; 5(1):vdad027.

54. **4.** Vossough A, Khalili N, Familiar AM, et al. Training and Comparison of nnU-Net and DeepMedic Methods for Autosegmentation of Pediatric Brain Tumors. *American Journal of Neuroradiology.* 2024.

55. **5.** Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences.* 2005; 102(43):15545-15550.

56. **6.** Fabregat A, Jupe S, Matthews L, et al. The reactome pathway knowledgebase. *Nucleic acids research.* 2018; 46(D1):D649-D655.

57. **7.** Liberzon A, Subramanian A, Pinchback R, Thorvaldsdóttir H, Tamayo P, Mesirov JP. Molecular signatures database (MSigDB) 3.0. *Bioinformatics.* 2011; 27(12):1739-1740.

58. **8.** Yu G, Wang L-G, Han Y, He Q-Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics: a journal of integrative biology.* 2012; 16(5):284-287.

59. **9.** Russo PS, Ferreira GR, Cardozo LE, et al. CEMiTool: a Bioconductor package for performing comprehensive modular co-expression analyses.

BMC bioinformatics. 2018; 19(1):1-13.

60. **10.** Rahmatallah Y, Emmert-Streib F, Glazko G. Gene Sets Net Correlations Analysis (GSNCA): a multivariate differential coexpression test for gene sets. *Bioinformatics.* 2014; 30(3):360-368.

61. **11.** Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology.* 2014; 15(12):1-21.

62. **12.** Korotkevich G, Sukhov V, Budin N, Shpak B, Artyomov MN, Sergushichev A. Fast gene set enrichment analysis. *BioRxiv.* 2016:060012.