Multiparametric MRI Along with Machine Learning Predicts Prognosis and Treatment Response in Pediatric Low-Grade Glioma

Anahita Fathi Kazerooni* [1,2,3,4], Adam Kraya* [1], Komal S. Rathi [1], Meen Chul Kim [1], Arastoo Vossough [1,5,6], Nastaran Khalili [1], Ariana M. Familiar [1], Deep Gandhi [1], Neda Khalili [1], Varun Kesherwani[1], Debanjan Haldar [1,7], Hannah Anderson [1,6], Run Jin [1], Aria Mahtabfar [7], Sina Bagheri [1,6], Yiran Guo [1], Qi Li [1], Xiaoyan Huang [1], Yuankun Zhu [1], Alex Sickler [1,3], Matthew R. Lueder [1,4,8], Saksham Phul [1], Mateusz Koptyra [1], Phillip B. Storm [1,3,4], Jeffrey B. Ware [6], Yuanquan Song [8,9], Christos Davatzikos [2,6], Jessica B. Foster [10,11], Sabine Mueller [12], Michael J. Fisher [10], Adam C. Resnick [1,4], Ali Nabavizadeh [1,6]

* These authors contributed equally

[1] *Center for Data-Driven Discovery in Biomedicine (D3b), The Children's Hospital of Philadelphia, Philadelphia, PA, USA*

[2] *AI^{2}D Center for AI and Data Science for Integrated Diagnostics, University of Pennsylvania, Philadelphia, PA, USA*

[3] *Department of Neurosurgery, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA*

[4] *Department of Neurosurgery, The Children's Hospital of Philadelphia, Philadelphia, PA, USA*

[5] *Division of Radiology, Children's Hospital of Philadelphia, Philadelphia, PA, USA*

[6] *Department of Radiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA*

[7] *Department of Neurosurgery, Thomas Jefferson University, Philadelphia, PA, USA*

[8] *Department of Pathology and Laboratory Medicine, The Children's Hospital of Philadelphia, Philadelphia, PA, USA*

[9] *Raymond G. Perelman Center for Cellular and Molecular Therapeutics, University of Pennsylvania, Philadelphia, PA, USA*

[10] *Division of Oncology, The Children's Hospital of Philadelphia, Philadelphia, PA, USA*

[11] *Department of Pediatrics, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA*

[12] *Department of Neurology and Pediatrics, University of California San Francisco, San Francisco, CA, USA*

Correspondence to Anahita Fathi Kazerooni (anahitaf@upenn.edu) or Ali Nabavizadeh (Ali.Nabavizadeh@pennmedicine.upenn.edu).

Abstract

Pediatric low-grade gliomas (pLGGs) exhibit heterogeneous prognoses and variable responses to treatment, leading to tumor progression and adverse outcomes in cases where complete resection is unachievable. Early prediction of treatment responsiveness and suitability for immunotherapy has the potential to improve clinical management and outcomes. Here, we present a radiogenomic analysis of pLGGs, integrating MRI and RNA sequencing data. We identify three immunologically distinct clusters, with one group characterized by increased immune activity and poorer prognosis, indicating potential benefit from immunotherapies. We develop a radiomic signature that predicts these immune profiles with over 80% accuracy. Furthermore, our clinicoradiomic model predicts progression-free survival and correlates with treatment response. We also identify genetic variants and transcriptomic pathways associated with progression risk, highlighting links to tumor growth and immune response. This radiogenomic study in pLGGs provides a framework for the identification of high-risk patients who may benefit from targeted therapies.

1 Introduction

Pediatric low-grade glioma (pLGG), classified as grade I or II tumors by the World Health Organization (WHO), constitutes the most common childhood brain tumors [1,2]. They represent roughly 30% of all pediatric central nervous system tumors and include a diverse spectrum of histologies, from glial to mixed glioneuronal entities [1,2]. The prognosis and treatment responses among patients with pLGG vary widely. Gross total resection offers the most favorable response in pLGG patients with 10-year progression-free survival (PFS) exceeding 85% [3]. Complete resection cannot be achieved for all tumors, especially for highly infiltrative or deep-seated tumors. Incompletely resected tumors often require additional therapy ranging from conventional chemotherapy to targeted inhibitors. Notably, for a significant portion of patients, the tumor progresses, resulting in a 10-year PFS rate under 50%, leading to concerns about long-term complications and an elevated mortality risk [3].

The potential of molecularly targeted therapeutics to transform pLGG treatment is increasingly recognized, as evidenced by the FDA approval and ongoing clinical trials of RAF and MEK inhibitors [4,5]. Nonetheless, a thorough understanding of the biological and molecular foundations of pLGG tumors, beyond merely identifying specific markers such as *BRAF* alterations, is crucial to the success of targeted treatments. This knowledge is also essential in preventing paradoxical tumor growth as seen with the sorafenib kinase inhibitor drug in *BRAF*-altered pLGGs [4]. Accordingly, with increasingly encouraging reports on the efficacy of immunotherapies in a number of solid pediatric brain tumors [6], it is vital to understand the immunological characteristics of brain tumors.

The tumor immune microenvironment (TIME) – composed of a diverse set of immune and stromal cells exhibiting complex interactions with each other and with tumor cells – influences the suppression or promotion of tumor growth. It plays a significant role in shaping the prognosis and affecting the response to treatment [7].

Radiomics has demonstrated potential in offering non-invasive, *in vivo* biomarkers of tumor molecular composition, which complement well-established clinical variables in supporting risk stratification by quantifying the underlying tumor heterogeneity and revealing the molecular underpinnings of the tumor via radiographic phenotypes (i.e., radiophenotypes) [8]. We propose that in pLGG, the use of radiomic analysis on treatment-naïve multiparametric MRI (mpMRI) sequences coupled with machine learning (ML) can reveal associations between radiophenotypes and the tumor's molecular composition, encompassing the TIME and prospective treatment targets. This approach extends the scope of analysis beyond the capabilities of molecular subtyping, which has traditionally been the focus of radiogenomic research in pLGGs [9-12]. By integrating imaging-derived phenotypes of pLGG with genotypic traits discerned from transcriptional analysis, our goal is to offer an in-depth characterization of pLGGs, contributing to risk stratification and identification of high-risk immunogenic patients who may benefit from novel precision therapies.

This study provides a comprehensive radiogenomic analysis through creating a model for characterization of pLGG tumor immune microenvironment (TIME) subgroups, developing a radioimmunomic signature for predicting these immunological subgroups, and generating a clinicoradiomic model for predicting pLGG progression-free survival,

assessing progression risk and treatment response. Additionally, an interpretability analysis was conducted to identify pathways associated with the predicted risk.

2 Results

2.1 Study Design and Cohort Description

The study design focuses on three main elements (Figure 1): (1) predicting pLGG immunological profiles, (2) predicting the immunological subgroups using a radioimmunomic signature, (3) developing a clinicoradiomic model to predict progression-free survival, along with an interpretability analysis of transcriptome pathways that are related to the risk of tumor progression. We analyzed retrospective data from a total of 545 patients (See Table 1 for detailed description) in the Children's Brain Tumor Network (CBTN [13]), encompassing 494 with multi-omic (RNA-seq and WGS) data and 201 with standard multiparametric MRI (mpMRI) scans. A subset of 150 had both WGS and mpMRI data. The imaging cohort for the clinicoradiomic risk stratification analysis, consisting of 201 patients, was randomly divided into two groups: a 'discovery' cohort with 160 patients for training and validation of the predictive models, and an 'unseen replication' cohort with 41 patients for independent testing of the trained models. For the radioimmunomic analysis, a supervised classification approach was used, which required the availability of both imaging and transcriptomic data. Maintaining the same 'discovery' and 'replication' cohort assignments, 150 subjects were found to have both imaging and transcriptomic data. Of these, 120 patients were included in the 'discovery' set, and 30 patients were included in the 'replication' set.

2.2 Transcriptome Enrichment Analysis Identifies Three Immunological Clusters

Clustering of 494 pLGGs with available RNA-sequencing data in our cohort revealed three distinct immunological groups based on immune cell infiltration-related gene expression, with immune cluster 1 (n = 189) showing intermediate inferred infiltration, immune cluster 2 (n = 164) showing high inferred infiltration, and immune cluster 3 (n = 141) showing the lowest inferred immune cell infiltration (Figure 2A). We observed a significant negative correlation between immune score and estimates of tumor purity derived from copy number (ABSOLUTE) [14], expression (ESTIMATE) [15], and methylation-based assessments (LUMP) [16] (Supplementary Figure 1).

Cluster 1 showed highest enrichment in stromal scores, eosinophils, and cancer-associated fibroblasts (CAFs) and lower levels of hematopoietic stem cells (HSCs), memory CD4+ T-cells, and CD4+ Th1 cells than clusters 2 and 3 (Figure 2B). Cluster 2 illustrated higher levels of common lymphoid progenitors, M2-polarized macrophages, and CD4+ Memory T-cells and the lowest enrichments for class-switched memory B-cells, eosinophils, T-regulatory cells (Tregs), and stromal scores compared to clusters 1 and 3 (Figure 2B). Cluster 3 exhibited higher expression of hematopoietic stem cells (HSCs), Tregs, and naïve CD8+ T-cells and lower levels of activated myeloid dendritic cells, monocytes, and M1-polarized macrophages in comparison with clusters 1 and 2 (Figure 2B). Analysis of variance (ANOVA) combined with Tukey's Honestly Significant Differences (HSD) test indicated that the composite immune score, an enrichment-based estimate of overall immune cell infiltration, was significantly higher in clusters 1

(mean immune score = 0.10) and 2 (mean immune score = 0.11) relative to cluster 3 (mean immune score = 0.038; adjusted p < 1e-7).

We also found that expression of the tumor inflammation signature (TIS), a clinically predictive expression-based measure of response to nivolumab and pembrolizumab [17], was statistically significantly different between clusters 1 and 2 (p = 2.6e-6), clusters 1 and 3 (p = 2.4e-8), and clusters 2 and 3 (p = 2.2e-16) (Figure 3 C & D). Cluster 2 revealed the highest TIS across all the immune groups, consistent with the independent measures of immune infiltration-related genes that we observed from xCell [18] with this group. Notably, tumor mutational burden (TMB), a clinical biomarker of response to immune checkpoint blockade [19,20], was significantly different between Cluster 1 (lower quartile = 0.17, median TMB = 0.32, upper quartile = 0.56) and Cluster 2 (lower quartile = 0.14, median TMB = 0.22, upper quartile = 0.31; p = 3.9e-5) and between Cluster 1 and Cluster 3 (lower quartile = 0.14, median TMB = 0.19, upper quartile = 0.33; p = 6.3e-5), but not between Clusters 2 and 3 (p = 0.75) (Figure 3E).

We then aimed to ascertain the relationship between immune groups and WHO 2021 brain tumor classifications of the CBTN cohort (Figure 3F, Supplementary Figure 4) [2]. To delineate specific xCell cluster and WHO tumor entities that deviated from expectation, we quantitatively determined significant group-level associations through a Poisson generalized linear model (GLM) and evaluated the Pearson residuals via a correspondence analysis. We found a highly significant, positive interaction (p < 0.001) effect between MAPK-pathway altered diffuse low-grade gliomas relative and clusters 1 and 2 (Figure 3F, Supplementary Figure 4), indicating over-representation of tumors

with a higher expression of immune infiltrate-related genes with this tumor entity. Pilocytic astrocytomas, often characterized by *KIAA1549::BRAF* fusion events, demonstrated a highly significant ($p < 0.001$) over-representation of cluster 2 tumors, characterized by the highest expression of infiltrate-related genes and the TIS. Gangliogliomas were significantly under-represented among cluster 2 tumors ($p = 0.012$; Figure 3F, Supplementary Figure 4). By contrast, cluster 3 tumors with lower expression of immune cell infiltrate-related genes were positively associated with MYB-altered diffuse astrocytomas, subependymal giant cell astrocytomas, and gangliogliomas (Figure 3F, Supplementary Figure 4). Collectively, these data suggest that tumor-cell intrinsic factors may be important mechanistic determinants of the tumor-immune phenotype in pLGGs, an observation that has also been noted in a number of other tumor types[21-23]. These data, along with our observations of associations between immune clusters with TIS and TMB, suggest that classifications defined in part by driver molecular aberrations may be more important and specific determinants with respect to the tumor-immune microenvironment and that total mutational count may play a secondary role given our observations between Clusters 2 and 3[19,20].

We then evaluated the association of the three immunological clusters with prognosis using survival endpoints progression-free survival (PFS) and overall survival (OS) (Figure 3 A&B; Supplementary Figure 2). A log-rank comparison across the three clusters revealed that the immune-high cluster 2 had the worst overall prognosis, followed by cluster 1 which illustrated intermediate infiltration-related gene expression, and finally cluster 3, which was predicted to have a low level of immune cell infiltration.

Notably, the cluster 2 subgroup had the highest expression of TIS and cluster 3 the lowest, and TIS has been shown to inversely associate with survival in adult low-grade gliomas consistent with our observations [24]. When modeling the effects of cluster when adjusted for age at diagnosis, reported sex, race, the anatomical region of the tumor, and the 2021 WHO classification of the tumor in a Cox regression model, we found significantly higher progression-related hazard for tumors belonging to cluster 2 ($p = 0.0081$, Figure 3B; Supplementary Table 3). The OS outcomes were not statistically different among the three patient groups by non-parametric Kaplan Meier analysis (Supplementary Figure 2). Further, PFS survival outcomes related to TIS, whether by log-rank comparisons across high/low groups or by Cox modeling adjusted for 2021 WHO classifications (adjustment in EFS model only due to convergence issues with OS modeling), showed that high TIS was associated with shorter PFS (Supplementary Figures 3A and 3B; log-rank $p = 0.023$ and Cox $p = 0.04$). OS models did not reveal a statistically significant difference across TIS groups (Supplementary Figures 3C and 3D).

We evaluated the relationship between immunological clusters and tumor locations, investigating if specific brain regions are inclined to exhibit distinct immune microenvironments. As the Sankey diagram in Supplementary Figure 5 presents, the immunological clusters are distributed fairly evenly across the cerebellar region, which is the most common tumor location across all three immunological clusters, followed by lobar tumor location (i.e., frontal, parietal, temporal, or occipital lobes). Suprasellar tumors were predominantly associated with immune-high clusters (1 and 2) and were

less common in the immune-low cluster 3. This pattern indicates that there is no significant predilection for specific brain regions for immunological profiles.

2.3 Radioimmunomic Signature: Differentiating Between Poor and Favorable Prognosis Immune Clusters

We assessed the potential of radiomic analysis in predicting pLGG immunological clusters, focusing on developing a radioimmunomic signature using machine learning classifiers. Since immune clusters 1 and 3 were not prognostically distinct (Figure 3A), we aimed to distinguish cluster 2 from clusters 1 and 3. Our analysis involved 150 patients with conventional MRI sequences, out of which 91 patients had additional ADC-map data. We trained radioimmunomic models on imaging features and age for prediction, testing them independently on the replication set. The conventional MRI-based radioimmunomic signature for cluster 2 versus clusters 1 and 3 showed an AUC of 0.77 | 0.74 and balanced accuracy of 76.8% | 86.0% in the discovery | replication sets. Adding ADC features to conventional MRI improved performance, with an AUC of 0.83 | 0.79 and balanced accuracy of 81.5% | 84.4% in the discovery | replication sets (Figure 4 A).

Our analysis further demonstrated different imaging characteristics within tumor subregions across the three immunological clusters (Figure 4 B). Immune cluster 2 exhibits more pixels with low enhancement in the ED tumor subregions, while cluster 3, characterized by lower immune cell expression and higher tumor purity (section 2.1.1 and Supplementary Figure 1), showed more pixels in the lower ADC range in the ED

region. This possibly indicates restricted diffusion and a higher tumor cell density rather than immune cell infiltration in peritumoral edema. A negative correlation between ADC values and tumor purity within the peritumoral edema has been reported in adult high-grade glioma [25].

2.4 Predictive Clinicoradiomic Model: Assessing Progression-Free Survival and Treatment Resistance Association

We hypothesized that a combination of clinical variables and radiomic features derived from pre-surgical/pre-treatment MRI sequences could predict tumor progression risk, thereby creating a non-invasive marker for the early prediction of progression risk. In our discovery cohort (n = 160), we developed a Cox proportional hazards model combining these variables. Ridge regularization was applied to radiomic variables, while clinical variables remained unpenalized (refer to Methods section 5.4). This model, tested on the replication set, aimed to predict patient risk scores (Figure 5A). In the discovery and replication sets, the clinicoradiomic model achieved Harrell's concordance indexes of 0.71 (95% CI = [0.63, 0.79]) and 0.77, Uno's concordance indexes of 0.72 (95% CI = [0.64, 0.80]) and 0.80, and integrated Brier scores (IBS) of 0.24 (95% CI = [0.19, 0.29]) and 0.16, respectively (Table 2). Supplementary Figure 6 shows hazard ratios for all variables in ascending order. Key clinical predictors include tumor resection extent, chemotherapy, age at diagnosis, and tumor location. Regularization selected 22 radiomic features, with 8 statistically significant. Patients were categorized into low, medium, and high-risk groups, with risk cutoffs applied

independently in the replication set. ANOVA tests yielded p = 3.12e-58 and p = 1.29e-54 across risk groups in the discovery and replication sets.

Our analysis revealed that a model based solely on clinical variables did not yield markedly different results from those of the clinicoradiomic model, as detailed in Table 2 and Supplementary Note 3. Nevertheless, given the notable predictive importance of total resection in pLGGs, we compared the efficacy of the trained clinicoradiomic and clinical-only models specifically in patients who did not undergo total/near-total resection. In this subset, the clinicoradiomic model demonstrated significantly enhanced performance relative to the clinical-only model, as evidenced by higher Harrell's concordance indexes (0.70 vs 0.55), Uno's concordance indexes (0.70 vs 0.57), and a lower IBS of 0.18 compared to 0.24 for the clinicoradiomic and clinical-only models, respectively (Table 2). This finding underscores the clinicoradiomic model's potential value in stratifying patients who are not adequately classified using only clinical variables.

We investigated the relationship between our predicted risk scores and treatment resistance, focusing on patients who showed progression (n = 61) within five years post-diagnosis. These patients were divided into two groups based on their treatment: those receiving minimal systemic therapy (chemotherapy or, rarely, radiotherapy post biopsy/surgery; n = 31), forming a treatment low-risk group, and those undergoing multiple treatments (n = 30), forming a treatment high-risk group. A Student's t-test revealed a significant difference in mean predicted risk scores between these groups (p

= 0.0010), with higher risk scores linked to patients less responsive to initial treatments, as shown in Figure 5B.

Our analysis of the clinicoradiomic risk groups in relation to treatment risk categories revealed a significant association, indicated by a Chi-squared test statistic of 13.69 and p = 0.0011. As Figure 5C shows, there is an over-representation of clinicoradiomic high-risk patients in the treatment high-risk group, and an under-representation in the treatment low-risk group. Only three clinicoradiomic low-risk patients showed progression, and they were all grouped into the treatment low-risk category. Most medium-risk clinicoradiomic patients fall into the treatment low-risk category. These findings imply that patients with higher clinicoradiomic risk scores may require multiple lines of treatment.

We explored correlations between clinicoradiomic risk categories and immunological clusters. Figure 5D illustrates this relationship, showing a rare occurrence of immune cluster 3 in the clinicoradiomic high-risk group, and a high occurrence of the prognostically worse immune cluster 2 in the clinicoradiomic high or medium-risk groups. Most of the tumors in immune cluster 1 were grouped in the medium and low-risk groups.

Our findings show that immune cluster 3, linked to a favorable prognosis, rarely appears in the predicted high-risk category, validating our clinicoradiomic model's ability to predict immunological factors related to progression risk. Predominantly, the clinicoradiomic high-risk group includes patients from immunological clusters 1 and 2, characterized by an inflamed yet possibly tumor-promoting immune environment.

Furthermore, we examined how molecular subtypes, as defined by Ryall et al [26], enhance the clinicoradiomic model's accuracy in predicting tumor progression risk. This analysis yielded Harrell's concordance indexes of 0.74 (95% CI = [0.65, 0.83]) and 0.69, Uno's concordance indexes of 0.74 (95% CI = [0.65, 0.84]) and 0.74, and IBS of 0.24 (95% CI = [0.14, 0.27]) and 0.18, for the discovery and the replication sets, respectively, as detailed in Supplementary Note 3, section 3.3.

Finally, considering the differences in progression-free survival probability between immunological clusters, we evaluated the performance of a model incorporating immunological clusters as a categorical variable (with three distinct values), along with clinical and radiomic variables. This model showed comparable performance to the clinicoradiomic model, achieving Harrell's concordance indexes of 0.71 (95% CI = [0.58, 0.83]) in the discovery set and 0.76 in the replication set, Uno's concordance indexes of 0.70 (95% CI = [0.60, 0.81]) in the discovery set and 0.78 in the replication set, and an IBS of 0.23 (95% CI = [0.16, 0.30]) in the discovery set and 0.14 in the replication set.

Given that immune cluster 2 was associated with a poorer prognosis compared to clusters 1 and 3, we trained a Cox model incorporating clinical and radiomic variables, along with a binary variable for immune clusters. In this model, immune cluster 2 was assigned one value, while clusters 1 and 3 were combined and assigned another value. As shown in Supplementary Table 8, this model's performance improved in the replication set compared to our clinicoradiomic model, with Harrell's concordance index of 0.80, Uno's concordance index of 0.83, and an IBS of 0.13. These results suggest that incorporating immunological clustering significantly enhances risk stratification in

Cox regression analysis, with immune clusters being a significant variable in these predictive models.

2.5 Biological Interpretability of Clinicoradiomic Risk Scores: Associations with Germline Variants and Transcriptomic Pathways

We then aimed to independently investigate whether genomic and transcriptomic characteristics may be associated with radiomically-defined risk of progression. We interrogated deleterious and likely deleterious germline variants present in the pLGG population through the criteria outline in Materials and Methods (Section 4.1.2.3), annotating for presence or absence of a germline variant among cancer genes and those involved in structural birth defects [27-30] (Supplementary Data 1). From a chi-square test of independence testing for the association of gene-specific germline variant counts across clinicoradiomic risk groups, we identified the synaptic nuclear envelope protein 1 (*SYNE1*) gene as significantly over-represented in the medium risk clinicoradiomic group (adjusted $p = 0.015$). Notably, *SYNE1* germline mutations have been implicated in cerebellar ataxia and somatic mutations of the gene are known to contribute to the development and progression of multiple cancers, among them adult glioblastoma [31-34].

We performed single-sample GSEA (ssGSEA) to derive per-sample enrichment scores for 1692 Reactome pathways spanning across all known cellular biological functions. Consistent with prior studies [23,35], we found that the *KIAA1549::BRAF* fusion subtype was linked to lower risk of progression. Among notable pathways negatively associated

with risk and cancer initiation/progression, we identified CD209+ dendritic cells, with an important role in initiating [36] and non-canonical NFKB signaling [37-39]. Furthermore, we found that up-regulated signaling of the eicosanoid family of lipid mediators, produced by poly-unsaturated fatty acids, were associated with higher risk, consistent with their known pro-angiogenic, proliferative, and inflammatory effects in glioma [40-42]. We identified several inflammatory pathways with known biological significance in glioma that were associated with greater risk. Interleukin-3, suppressing anti-tumor immune responses via activation of PDL1 pathway [43], and Interleukin-5 signaling that promotes cancer metastasis by remodeling the TIME [44], were associated with greater risk of progression. Type I interferon signaling, previously found to promote a mesenchymal phenotype and poor prognosis in glioma [45]. Additional pathways that remodel the inflammatory phenotype and promote tumor growth or proliferation, and were associated with higher risk, included alpha protein kinase 1 [46], TRAF6-mediated NFKB activation [47], and oncogene-induced senescence pathways [48].

3 Discussion

In this study, we undertook a comprehensive radiogenomic analysis of pLGGs using both supervised and unsupervised ML techniques. Our primary objective is to pave the path for curing pLGGs while reducing the need for invasive tumor resection or systemic treatments, thereby reducing related morbidities in children. We focused on identifying intrinsic imaging and molecular characteristics of pLGG tumors, exploring their phenotype-genotype relationships from various analytical perspectives for distinct

clinical applications. This method also addresses a major hurdle in ML: the lack of biological or clinical interpretability which hinders its integration into treatment decisions.

The analysis of immunological clusters revealed their association with clinical outcomes, wherein immune-high cluster 2 was associated with worst progression-free survival rates, likely linked to enrichment of M2-polarized macrophages, which suppress local tumor immunity and mast cells reshaping the TIME to promote angiogenesis [49,50]. Conversely, memory B-cells that favor anti-tumor T-cell mediated immunity in glioma [37] were relatively low in cluster 2 and higher in prognostically-favorable cluster 1.

Immune cluster 1, with more favorable prognosis, showed relatively higher inferred levels of eosinophils, also observed in independent LGG cohorts [51]. In cluster 3, which showed the longest median overall survival of all immune groups, we detected signatures related to NK and CD8+ cell types, both reported to confer more favorable prognosis in glioblastoma [52-55]. This suggests that profiling immune cell types in pLGG may provide insights with respect to disease prognosis.

These findings also suggest the potential benefit of immune clustering in determining response to and stratifying patients for checkpoint inhibitor therapy. *BRAF* is vital in the *MAPK* signaling pathway, influencing various cellular functions like proliferation and angiogenesis [56]. Although *BRAF* inhibitors are potential treatments in pLGGs [57,58], there is a prevalent issue of therapeutic resistance [59] and serious side effects like secondary neoplasms [60]. As such, exploring alternative treatments like immunotherapy, either on its own or combined with targeted therapies, may be beneficial for pLGGs with *BRAF* mutations [61]. In our study, we discovered that *BRAF*-altered pLGGs exhibit increased

immune cell infiltration-related gene expression. It has been shown that the *BRAF V600E* mutation may boost immunogenicity in LGG patients, leading to the development of *BRAF V600E*-specific T-cells, which can then be targeted with T-cell therapy [62]. However, another study focusing on pLGGs found no link between PD-1 expression levels and the *BRAF V600E* mutation status [63]. As we continue to make advancements in molecular research, further investigations will be essential to pinpoint the optimal immunotherapeutic strategies for pLGG patients.

In our research, a strong correlation was found between higher TIS and immunological groups in pLGGs, whereas TMB was relatively high in only one cluster with relatively inferred immune cell content (Cluster 1). This is consistent with a previous study on angiosarcoma patients where TIS was linked to specific immune clusters, including neutrophils, macrophages, and PD-L1+ cells [64]. Furthermore, a breast cancer study found TIS scores closely aligned with tumor-infiltrating lymphocyte numbers [65]. Pan-cancer studies have also recognized TIS as a consistent predictor of immunotherapy response across multiple cancers [17].

Wang et al. [66] previously reported the immunological classification of pediatric gliomas, including both low- and high-grade tumors, based on RNA-sequencing. They identified three primary immunological subgroups, with the immune-hot group showing a higher presence of M2-polarized macrophages and CD4+ Memory T-cells, which aligns with our findings. However, unlike their study, we did not observe any differences in OS among the three immunological subgroups. This discrepancy could primarily be due to Wang et al.'s inclusion of high-grade tumors in their analysis. In contrast, pLGGs usually

have a prolonged OS, leading to a higher proportion of censored data in our study. In our research, we advanced the methodology by developing an imaging-based radioimmunomic signature that predicts immunological clusters. This signature acts as a surrogate biomarker for immunological profiles in situations where RNA-sequencing is unfeasible, such as when the tumor location impedes surgical resection, or in settings where RNA-sequencing and transcriptomic analysis are not possible due to limited resources.

The ability of our imaging-based ML model to distinguish the 'immune-hot' cluster 2 from immune clusters 1 and 3, each harboring distinct cellular profiles, may prove clinically beneficial when it comes to predicting the efficacy of immune-based therapies in pLGG upfront. Radiomics has been used for predicting immune phenotypes [67] and specific immune cell populations [68-70] in adult glioblastoma. Nonetheless, our investigation stands as the pioneering effort to create imaging-based ML models capable of predicting immune clusters in pediatric brain tumors. Our study highlights the value of integrating ADC-derived features that provide quantitative assessments of pattern of diffusivity of water molecules in the TIME. The reports on the positive or negative correlation of ADC values and tumor microenvironment markers remain contradictory [69-71]. Moreover, a radiomic study based solely on ADC maps did not outperform a model constructed using post-contrast T1-weighted MRI sequences, indicating the complexity of identifying the most informative imaging features for such predictive tasks [67]. While studies in the adult brain tumor domain have pointed to the improved accuracy of radiogenomic classifications after the inclusion of features from advanced MRI

sequences such as ADC map [8,72], there remains a paucity of research that directly correlates ADC values with the TIME in brain tumors [70,71].

We sought to determine the predictive value of clinically significant variables, known for their prognostic impact in pLGG, and radiomic features from pre-surgical MRIs for assessing tumor progression risk. The high reproducibility of our clinicoradiomic model, evident with concordance indexes of 0.71 and 0.77 in the discovery and replication sets, respectively, indicates its potential for future application in unseen patient cohorts, including prospective studies. Consistent with existing literature [73,74], the extent of tumor resection was a key determinant of progression-free survival. In our study, among 61 progressed tumors, only 7 had undergone total or near-total resection, with progression-free intervals ranging from 222 to 921 days and a median of 660 days post-surgery.

When comparing our clinicoradiomic risk groups with treatment risk categories, we found that most treatment high-risk tumors were also identified as clinicoradiomic high-risk tumors, while a few were in the medium-risk category. Most treatment low-risk tumors corresponded to our low to medium-risk clinicoradiomic groups. Notably, the treatment low-risk tumors categorized as high-risk by our clinicoradiomic model still progressed within five years, (progression-free survival range 4.8 - 131.5, median, 19.2 months), suggesting accuracy of our model in predicting progression-free survival and progression risk. These findings underscore our model's potential to guide treatment strategies, suggesting consideration for a targeted therapy approach or enrollment onto a clinical trial upfront, rather than at recurrence, for 'high-risk' tumors and more conservative, monitoring-focused approaches for 'low-risk' cases. The 'medium-risk'

category could be considered for a proactive treatment approach [3], although it currently lacks a clear clinical definition in pLGG tumors, which complicates understanding its clinical implications. However, our findings indicate an over-representation of *SYNE1* gene, linked to cerebellar ataxia and tumor progression, in the medium-risk clinicoradiomic group. Considering the long-term morbidities associated with pLGG tumors, including ataxia, children in this risk category may be predisposed to the development of progressive ataxia [75], underscoring the importance of upfront risk stratification to either prevent the onset or enable early management of these side effects.

Biological interpretation of our clinicoradiomic model via radiogenomic analysis showed that the predicted risk scores were associated with molecular pathways known to influence tumor progression which may provide insights for clinical interventions. *NFKB* signaling, a pathway known for the maintenance of B-cells and its role in T-cell driven anti-tumor immune responses in glioblastoma [46-48], was negatively associated with risk. Pathways related to the balance of fatty acid synthesis and oxidation, as a metabolic switch to promote lipogenesis and lipid droplet formation – a well-established tumor-promoting mechanism across cancers particularly high-grade gliomas [42,43] – were found to relate to elevated risk. IL-3, which suppresses anti-tumor immune responses via activation of the PDL1 pathway [43], and IL-5 signaling, shown to promote cancer metastasis by remodeling the immune microenvironment [44], were also linked to higher risk.

The existing literature on radiomic analysis for pediatric brain low-grade glioma is limited, primarily concentrating on the prediction of single genetic mutations or alterations, such as *BRAF-V600E* mutation and *BRAF-KIAA1549* fusion [9-12]. In light of the 2021 WHO classification of tumors of the central nervous system (CNS), 5th edition (WHO CNS 5) [2], which recommends the inclusion of driver molecular alterations in standard diagnostic processes, merely predicting a single mutation or fusion using imaging might not yield substantial clinical benefits. Our study offers a comprehensive radiogenomic analysis that integrates radiomics with transcriptomic pathways, extending beyond the predictive modeling of single molecular drivers and addressing a research gap in pediatric and adult brain tumor studies.

The current study encountered challenges due to the limited availability of imaging and genomic data (WGS or RNA-sequencing), compounded by the rarity of pediatric brain tumors. We took rigorous precautions in training our machine learning models by splitting the available data into a “discovery” set for training and validation, using a thorough nested cross-validation approach. Additionally, we employed a fully independent “replication” set, unseen during model training, to assess the generalizability of our models to new data. Despite these measures, our imaging cohort was relatively small and included data from only a single institution, highlighting the need for further external validation to test its generalizability.

The cohort comprised a limited number of patients with *NF1* disease, non-pilocytic astrocytoma histologies, and was confined to specific tumor sites. The time of progression was determined by reviewing clinical notes on treatment changes and MRI

scans when progression was suspected, though this method may not be entirely objective. Future studies could improve accuracy by employing volumetric analysis, although its advantage over the traditional bidimensional method is not yet proven.

Our analysis comprised radiomic, genomic, and transcriptomic data, whereas other important layers like methylation profiling, metabolomics, proteomics, and pathomics were not available for most of the patients. To gain a deeper understanding of the complex interplay between these data layers at micro and macro scales, and how they influence the clinical presentation and treatment response of tumors, future studies incorporating a broader spectrum of data could further inform precision medicine approaches for individual patients. Additionally, due to unavailable tissue, we were unable to validate our transcriptomic findings through immunohistochemical staining, suggesting a need for further research to confirm cell enrichments in immunological subgroups.

Acknowledging these limitations, our analysis sets the stage for the development of more personalized and less invasive treatment strategies for pLGGs in future research. Specifically:

- The discovery of the radioimmunomic signature provides a non-invasive technique to identify immune-high pLGG tumors, with elevated TIS levels and typically a poorer prognosis, which may be candidates for immunotherapy. This approach can aid in the early stratification of patients, allowing for their inclusion in immunotherapy trials.

- The clinicoradiomic predictive model can serve as an effective marker for assessing patient's risk of progression and predicting treatment response. This model can be applied as a pre-interventional assessment tool, enabling consideration of different scenarios, such as varying degrees of surgical resection and chemotherapy use. Its predictive capability allows for the simulation of diverse clinical outcomes based on hypothetical treatment strategies prior to making patient management decisions. Moreover, the predictions of the clinicoradiomic model can be updated and recalibrated after each intervention, or as new data about tumor molecular subtypes emerge post-resection. This iterative refinement ensures that the prognostic evaluations stay accurate and applicable to the patient's changing clinical condition. Future enhancements to the model, including the incorporation of multi-omic data, will further refine its accuracy and utility. Integrating this predictive model into clinical practice will provide valuable guidance in treatment planning, thereby enhancing the management and care of patients with pLGGs.

Overall, by leveraging the insights from our analysis, a more patient-centric approach can be adopted, tailoring treatments to the specific characteristics and needs of each patient, thus enhancing the overall quality and effectiveness of care for patients with pLGGs, aligning closely with the evolving field of personalized medicine.

4 Methods

In this study, we utilized retrospectively collected data from the Children's Brain Tumor Network (CBTN) repository (cbtn.org)[13]. The repository hosts a comprehensive

collection of specimens, along with longitudinal clinical and imaging data, designed to support the sharing of de-identified samples and information for ongoing research. The study adhered to HIPAA regulations and received IRB approval from the Children's Hospital of Philadelphia (CHOP) under the CBTN protocol, with informed consent obtained for patient participation.

4.1 Data

4.1.1 Overview of the Data

Data from patients diagnosed histopathologically with *de novo* pediatric low-grade glioma (pLGG). RNA-sequencing data was available for 494 subjects from a combination of CBTN patient data and patient data available in the OpenPedCan repository (v12 data release), which is a collection of genomic and molecular data available on multiple sources, such as KidsFirst (https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs001436.v1.p1), Therapeutically Applicable Research to Generate Effective Treatments (TARGET; https://www.cancer.gov/ccg/research/genome-sequencing/target), OpenPedCan (https://github.com/PediatricOpenTargets/OpenPedCan-analysis) [76], and The Cancer Genome Atlas (TCGA; https://www.cancer.gov/about-nci/organization/ccg/research/structural-genomics/tcga). Subjects had molecular subtype information available through the OpenPedCan database and through the application of an in-house molecular subtyping pipeline on additional cases consented

on CBTN's research protocol after the v12 OpenPedCan release [76]. Parents or legal guardians provide consent utilizing CBTN regulatory documentation for pediatric patients, and children also assent if they are age 7 or above and are cognitively able to. Race is self-reported and was extracted from the electronic medical records of the participants. Molecular subtypes were defined based on those reported by Ryall et al. [26] pLGG WHO 2021 tumor classifications were defined based on the reported pathological diagnosis of the tumor and in part based on the presence/absence of driver molecular aberrations using OpenPedCan consensus variant calling data. Biological sex was determined through samtools (idxstats function) analysis of the X- and Y-chromosome sequencing reads [77].

Treatment-naïve 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), acquired between 2006 and 2018, from 258 subjects with pLGG were obtained. Diffusion weighted imaging (DWI) and scanner-generated apparent diffusion coefficient (ADC) maps were collected for 153 subjects. Patients were excluded if: (1) age at the time of imaging was older than 18 years, (2) the images were not treatment-naïve, i.e., were not acquired prior to the initial surgery or treatment, (3) the primary site was not brain, (4) leptomeningeal dissemination was reported, (5) not all four standard mpMRI sequences were available, or (6) the images were determined to be of low-quality. A final cohort of 201 subjects with available standard mpMRI (91 with ADC-maps) was included in the imaging studies. In cases where multiple imaging sessions were available for the selected patients before their initial

treatment or surgery, we chose the session that was closest in time to the intervention. The time interval between these sessions and the intervention for our cohort ranged from 1 to 370 days, with a median interval of 3 days.

For the patients in the final cohort, we reviewed the clinical notes for the patients that had progressed. For the subjects that had tumor progression, we recorded the number of radiotherapy or chemotherapy treatments the patients had received over the course of the study within the five years since the initial diagnosis. A treatment change was classified as a new treatment only in cases of treatment failure, not when a change in treatment method was required due to toxicity. Supplementary Tables 1-2, and Supplementary Data 2 provide details on clinical characteristics, MRI scanner information, and the available imaging acquisition parameters, respectively, in the imaging cohort.

4.1.2 Molecular Data Analysis

4.1.2.1 Enrichment analysis and clustering of immune and stromal cell types

To identify the enrichment of immune and stromal cell types and pathways, xCell [18], which is a cell type enrichment analysis from gene expression data, was applied on our pLGG transcriptomic data. We utilized transcripts per million (TPM) generated through a harmonized STAR-RSEM RNA-sequencing pipeline developed by the Gabriella Miller KidsFirst Data Resource Center as part of its sequencing and benchmarking initiative (https://kidsfirstdrc.org/). This pipeline incorporates Ensembl's GENCODE version 39 [78] for updated gene-level annotations. We leveraged the Immunedeconv R package [79]

specifying default parameters and xCell as the method of interest. Tumor purity was inferred using three different approaches, including estimates of tumor purity derived from copy number (ABSOLUTE) [14], expression (ESTIMATE) [15], and methylation-based assessments (LUMP) [16] (Supplementary Figure 1). Kruskal-Wallis's test was performed to determine the differences in tumor inflammation signature (TIS) and tumor mutational burden (TMB) levels among the immune clusters.

4.1.2.2 Enrichment analysis of molecular pathways

To identify differentially expressed biological pathways across clinicoradiomic risk groups, we filtered gene expression data for the 'protein coding' gene type. We then provided expected counts as input to DeSeq2 [80] for differential gene expression analysis, specifying clinicoradiomic risk groups as the factor variable in the model's design. Lists of differentially expressed genes derived from pairwise comparisons across all three imaging groups were ranked and used as input for pre-ranked GSEA [81] to identify differentially expressed biological pathways. Pathway annotations were derived from Reactome using the Molecular Signatures Database v2023.1.Hs [82].

4.1.2.3 Germline Variant Analysis

To identify deleterious/likely deleterious germline variants in a list of cancer predisposition genes, we applied the following criteria to filter out single nucleotide variants (SNVs) and small insertions/deletions (indels): a) minor allele frequency in gnomAD (v2.1.1 [83]) and TOPMed (data freeze 7 [84]) is less than or equal to 0.0001; b) at least half of 21 prediction scores as listed in Supplementary Data 1 of the dbNSFP

database (v4.2 [85,86]) label a variant as Deleterious, Detrimental, Damaging or Medium/High Impact; c) for quality control purposes we require that in heterozygous calls, alternate allele depth is between 25% and 75% of all sequencing depth of a variant site. We also checked if variants are cataloged in either ClinVar (release 20240331 [87]) or Human Gene Mutation Database (HGMD; Professional Version 2024Q1 [88]). Variants with membership in ClinVar/HGMD are considered "deleterious", while others obtained through the germline variant filtering process described above are considered "likely deleterious".

4.1.3 Imaging Data Processing and Analysis

In our study, image standardization was performed prior to radiomic feature extraction. For each patient, all MRI scans were first re-oriented to left-posterior-superior (LPS) coordinate system. Subsequently, T1w, T2w, T2-FLAIR, and ADC-maps (when available) were co-registered with their corresponding T1w-Gd sequence. All images were then resampled to an isotropic resolution of 1 mm^3 based on anatomical SRI24 atlas using the Greedy algorithm, in Cancer Imaging Phenomics Toolkit open-source software v.1.8.1 (CaPTk, https://www.cbica.upenn.edu/captk). All preprocessed images were then skull-stripped using our dedicated brain tissue extraction tool [89]. The skull-stripped images were then normalized to the intensity range of [0, 255] after removing after removing the outlier pixels that did not fall into 99.9% percentile of the image histogram. Brain tumor segmentation was performed using our in-house automatic pediatric brain tumor segmentation tool (https://github.com/d3b-center/peds-brain-auto-seg-public) [89], and manual revisions were made when necessary. This tool generates

segmentation of tumor subregions, including the enhancing tumor, nonenhancing tumor, cyst, and edema. A combination of all subregions is used to generate a whole lesion segmentation (indicated in Figure 1).

Radiomic features (n = 881) of shape, volume, intensity, first-order histogram (10 bins), 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 features were extracted after overlaying the WT mask on the mpMR images.

The data was randomly split into a “discovery” (n = 160) and an unseen “replication” (n = 41) cohort; since some tumor locations, such as basal ganglia, were rare in our cohort, we kept those subjects in the discovery set. We have used the same cohort split across all experiments throughout this study; however, some of the experiments had reduced sample sizes when intersected with transcriptomic data. Prior to each ML analyses, pairwise Pearson’s correlation was performed on the extracted radiomic features for the subjects in the discovery cohort, to remove one of the highly correlated pair of features ($r > 0.90$), reducing the dimension to 438. The features were then normalized using z-scoring approach. For categorical variables, including sex, *NF1* disease, tumor location, extent of tumor resection, chemotherapy, and radiotherapy, count encoding was performed. Age at diagnosis, as a continuous variable, was normalized using z-scoring method. The mean and standard deviation calculated for the continuous features in the discovery set were used to normalize the corresponding features in the replication set.

Histograms illustrating the most indicative features were produced for different groups of tumors in each of the experiments, to elucidate the inherent biological processes associated with specific genetic alterations. These histograms display the frequency (y-axis) of a particular feature value (x-axis), collated from the entire patient dataset.

4.2 Imaging-Based Prediction of TIME Subgroups via Radioimmunomic Signature

For this analysis, immunological subgroups of pLGG were determined using unsupervised clustering based on enrichment of immune and stromal cells (obtained from xCell), followed by application of supervised approach based on radiomic features for classification of the immune clusters (Figure 1 shows the workflow for this analysis).

4.2.1 Characterization of TIME Subgroups

We leveraged the xCell enrichment scores derived as described above as input for clustering. Clustering was carried out using an internally developed R package that comparatively and quantitatively assesses clustering quality across dissimilarity-based [90], model-based [91], joint dimensionality reduction-based [92], and dbscan [93] clustering methods. We leveraged the fpc R package (version 2.2-13; Hennig C, 2024, licensed under GPL-2) to derive a quantitative composite score using multiple metrics of clustering quality, selecting the highest scoring method based on cluster compactness and separation. The v-test, which is a measure of the difference in means across a given cluster relative to the remaining sample set scaled by the variance [94], was leveraged for identifying cluster-specific immune cell subsets. Heatmap plots were

generated using the pheatmap R package (version 1.0.12; Kolde R, 2024, licensed under GPL-2), violin plots using ggplot2 R package (version 3.5.1; Wickham H, et al., 2024, licensed under MIT) [95], and radar plots using the fmsb R package (version 0.7.6; Nakazawa M, 2024, licensed under GPL-2).

4.2.2 Radioimmunomic Signature

To generate a radiomic signature of immunological profiles, we used the intersection of our imaging and transcriptomic cohorts, resulting in a total of 150 subjects (120 in the discovery, 30 in the replication sets) with all four conventional MRI sequences (T1w, T1w-Gd, T2w, T2-FLAIR) and 91 subjects (72 in the discovery, 19 in the replication sets) when ADC-map was also available. On a feature set including the radiomic features (n = 438 when the four conventional MRI sequences were available and n = 493 when ADC was also included) and age at diagnosis, support vector machines (SVM), wrapped with Maximum Relevance Minimum Redundancy (MRMR) feature selection, was applied to classify the subjects in the discovery set into three immunological groups. ML classifier was trained using a nested cross-validation (CV) approach, with five folds for the inner loop for the purpose of feature selection and ten folds for the outer loop to tune the hyperparameters and optimize the model. This nested CV approach is intended to avoid data overfitting and improve generalizability to the unseen replication cohort. To reduce overfitting to the training data, we used a linear kernel for SVM. Three binary classifiers for each of the immunological groups were trained in a one-versus-the-rest approach. This implementation was carried out in MATLAB 2023a (MathWorks Inc). The generated ML models were then independently

applied to the replication set to evaluate the performance via area under the receiver operating characteristic (ROC) curve (AUC) and balanced accuracy.

4.3 Imaging-Based Prediction of Risk of Tumor Progression

Patient demographics and clinical variables considered to play an important role in determining the risk of the patient for tumor progression, i.e., age, sex, extent of tumor resection, treatment method (radiotherapy, chemotherapy), tumor location, and germline Neurofibromatosis (*NF1*) cancer predisposition were included as input variables. A baseline Cox-PH model only based on clinical variables was generated to study additional value of radiomics to these clinically available features. The data (n = 201) was split into a discovery (n = 160) and replication (n = 41) sets. The discovery set was used for training a Cox-PH model using a stratified nested cross-validation (10 folds for both inner and outer nests) approach. The trained model was applied on the "unseen" replication set. Consequently, we generated another Cox-PH model with ElasticNet penalization using a combination of clinical and radiomic variables. The penalization was set to zero for clinical variables but applied for radiomic features to decrease the dimensionality of the feature space. The performances of survival models were evaluated based on three metrics were employed: 1) Harrell's concordance index (hereinafter referred to as "HAR"), 2) Uno's concordance index (hereinafter referred to as "UNO"), and 3) the Integrated Brier Score (hereinafter referred to as "IBS"), which provides an overall calculation of the model performance at all available times. The patients were then stratified into the groups of low, medium, and high risk based on their predicted risk in the discovery set. To achieve this, we fitted the entire discovery cohort

against the survival model to obtain risk scores. Based on the distribution of risk scores calculated by the model, the patients were grouped into three categories: 1) High (>=75%), 2) Medium (>=25%; <75%), and 3) Low (< 25%).

4.4 Statistical Analysis

Tumor mutational burden (TMB) and tumor inflammation signature (TIS) were calculated from the whole genome sequencing (WGS) data using the R package "maftools" (version 2.22.0; Mayakonda A, et al., 2024, licensed under MIT) based on VarScan method [96]. Molecular Signatures Database (MSigDB, http://www.gsea-msigdb.org/gsea/index.jsp) was used to extract immune or inflammation gene sets and differential expression analysis was performed between tumor and normal tissues. Analysis of correlation between the categorical variables from the immune clusters and pLGG 2021 WHO classifications was carried out using a Poisson GLM in combination with a correspondence analysis of the Pearson residuals. Kruskal-Wallis and one-way analysis of variance (ANOVA) tests were performed to evaluate the statistical difference in the mean values of TMB and TIS, respectively, across the three different immune groups. Student's t-test was applied to calculate the difference between the mean values of the groups of patients that had progressed and had no or one treatment and more than one treatment. For evaluation of the performance of the multivariate classification approach, area under the receiver operating characteristic (ROC) curve (AUC) along with 95% confidence interval (95% CI) were calculated.

4.5 Data Availability

The processed genomic, transcriptomic, clinical, and radiomic data used in this study are available at https://github.com/d3b-center/pLGG-immune-clinicoradiomics, DOI: 10.5281/zenodo.13942516 [97]. The data generated for the figures in this manuscript are provided in the Supplementary Information/Source Data file. De-identified genomic, transcriptomic, and clinical data, including patient baseline characteristics, molecular subtypes, and outcomes, were obtained from publicly available resources. These datasets can be accessed through the dbGaP study under accession number phs002517.v4.p2 [https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs002517] and the KidsFirst Data Resource Portal (https://kidsfirstdrc.org/) under study ID phs002517. De-identified imaging data for patients, as identified by Subject IDs in our GitHub repository, can be obtained through the Children's Brain Tumor Network (CBTN) by submitting a request via the CBTN request form at cbtn.org.

4.6 Code Availability

The codes for analyzing processed clinical, genomic, transcriptomic, and radiomic data in relation to patient clinical characteristics is available under an Apache License 2.0 on GitHub: https://github.com/d3b-center/pLGG-immune-clinicoradiomics,
DOI: 10.5281/zenodo.13942516 [97].

All image processing tools used in this study are freely available for public use:

- Automated image processing and feature extraction codes are available at https://github.com/d3b-center/peds-brain-seg-pipeline-public (under BSD 2-Clause license). Portions of these codes are adapted from CaPTk (https://cbica.github.io/CaPTk), which is licensed under the CBICA Software License (https://www.med.upenn.edu/cbica/software-agreement.html). The original license information and copyright statements have been retained in the source files where applicable. The original CaPTk codes can be accessed at https://cbica.github.io/CaPTk.

- Tumor segmentation: https://github.com/d3b-center/peds-brain-auto-seg-public (under BSD 2-Clause license)

- Skull-stripping: https://github.com/d3b-center/peds-brain-auto-skull-strip (under BSD 2-Clause license)

5 References

1 Ostrom, Q. T. *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* **24**, iii1-iii38 (2022).

2 Louis, D. N. *et al.* The 2021 WHO classification of tumors of the central nervous system: a summary. *Neuro-oncology* **23**, 1231-1251 (2021).

3 Ryall, S., Tabori, U. & Hawkins, C. Pediatric low-grade glioma in the era of molecular diagnostics. *Acta neuropathologica communications* **8**, 1-22 (2020).

4 Manoharan, N., Liu, K. X., Mueller, S., Haas-Kogan, D. A. & Bandopadhayay, P. Pediatric low-grade glioma: Targeted therapeutics and clinical trials in the molecular era. *Neoplasia* **36**, 100857 (2023).

5 Jones, D. T. *et al.* Pediatric low-grade gliomas: next biologically driven steps. *Neuro-oncology* **20**, 160-173 (2018).

6 Khalili, N. *et al.* Radiomics for characterization of the glioma immune microenvironment. *NPJ Precision Oncology* **7**, 59 (2023).

7 Quail, D. F. & Joyce, J. A. The microenvironmental landscape of brain tumors. *Cancer cell* **31**, 326-341 (2017).
8 Fathi Kazerooni, A., Bakas, S., Saligheh Rad, H. & Davatzikos, C. Imaging signatures of glioblastoma molecular characteristics: a radiogenomics review. *Journal of Magnetic Resonance Imaging* **52**, 54-69 (2020).
9 Tak, D. *et al.* Noninvasive molecular subtyping of pediatric low-grade glioma with self-supervised transfer learning. *medRxiv* (2023).
10 Trasolini, A. *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* **43**, 1196-1201 (2022).
11 Wagner, M. *et al.* Radiomics of pediatric low-grade gliomas: toward a pretherapeutic differentiation of BRAF-mutated and BRAF-fused tumors. *American Journal of Neuroradiology* **42**, 759-765 (2021).
12 Kudus, K. *et al.* Increased confidence of radiomics facilitating pretherapeutic differentiation of BRAF-altered pediatric low-grade glioma. *European Radiology*, 1-10 (2023).
13 Lilly, J. V. *et al.* The children's brain tumor network (CBTN)-Accelerating research in pediatric central nervous system tumors through collaboration and open science. *Neoplasia* **35**, 100846 (2023).
14 Carter, S. L. *et al.* Absolute quantification of somatic DNA alterations in human cancer. *Nature biotechnology* **30**, 413-421 (2012).
15 Yoshihara, K. *et al.* Inferring tumour purity and stromal and immune cell admixture from expression data. *Nature communications* **4**, 2612 (2013).
16 Aran, D., Sirota, M. & Butte, A. J. Systematic pan-cancer analysis of tumour purity. *Nature communications* **6**, 8971 (2015).
17 Damotte, D. *et al.* The tumor inflammation signature (TIS) is associated with anti-PD-1 treatment benefit in the CERTIM pan-cancer cohort. *Journal of Translational Medicine* **17**, 1-10 (2019).
18 Aran, D., Hu, Z. & Butte, A. J. xCell: digitally portraying the tissue cellular heterogeneity landscape. *Genome Biology* **18**, 1-14 (2017).
19 Marabelle, A. *et al.* Association of tumour mutational burden with outcomes in patients with advanced solid tumours treated with pembrolizumab: prospective biomarker analysis of the multicohort, open-label, phase 2 KEYNOTE-158 study. *The Lancet Oncology* **21**, 1353-1365 (2020).
20 Marcus, L. *et al.* FDA approval summary: pembrolizumab for the treatment of tumor mutational burden–high solid tumors. *Clinical Cancer Research* **27**, 4685-4689 (2021).
21 Kraya, A. A. *et al.* PTEN loss and BRCA1 promoter Hypermethylation negatively predict for immunogenicity in BRCA-Deficient ovarian Cancer. *JCO Precision Oncology* **6**, e2100159 (2022).
22 Kraya, A. A. *et al.* Genomic Signatures Predict the Immunogenicity of BRCA-Deficient Breast CancerImmunogenetic Signatures of BRCA1/2 Breast Cancer. *Clinical Cancer Research* **25**, 4363-4374 (2019).

23 Li, J. *et al.* Tumor cell-intrinsic factors underlie heterogeneity of immune cell infiltration and response to immunotherapy. *Immunity* **49**, 178-193. e177 (2018).
24 Danaher, P. *et al.* Pan-cancer adaptive immune resistance as defined by the Tumor Inflammation Signature (TIS): results from The Cancer Genome Atlas (TCGA). *Journal for immunotherapy of cancer* **6**, 1-17 (2018).
25 Zhou, Q. *et al.* Correlation analysis of tumor purity with clinicopathological, molecular, and imaging features in high-grade gliomas. *Neurosurgical Review* **45**, 3699-3708 (2022).
26 Ryall, S. *et al.* Integrated molecular and clinical analysis of 1,000 pediatric low-grade gliomas. *Cancer cell* **37**, 569-583. e565 (2020).
27 Firth, H. V. *et al.* DECIPHER: database of chromosomal imbalance and phenotype in humans using ensembl resources. *The American Journal of Human Genetics* **84**, 524-533 (2009).
28 Tate, J. G. *et al.* COSMIC: the catalogue of somatic mutations in cancer. *Nucleic acids research* **47**, D941-D947 (2019).
29 Chakravarty, D. *et al.* OncoKB: a precision oncology knowledge base. *JCO precision oncology* **1**, 1-16 (2017).
30 Krysiak, K. *et al.* CIViCdb 2022: evolution of an open-access cancer variant interpretation knowledgebase. *Nucleic acids research* **51**, D1230-D1241 (2023).
31 Synofzik, M. *et al.* SYNE1 ataxia is a common recessive ataxia with major non-cerebellar features: a large multi-centre study. *Brain* **139**, 1378-1393 (2016).
32 Masica, D. L. & Karchin, R. Correlation of somatic mutation and expression identifies genes important in human glioblastoma progression and survival. *Cancer research* **71**, 4550-4561 (2011).
33 Chu, Y.-D. *et al.* SYNE1 exonic variant rs9479297 contributes to concurrent hepatocellular and transitional cell carcinoma double primary cancer. *Biomedicines* **9**, 1819 (2021).
34 Urbini, M. *et al.* Whole exome sequencing uncovers germline variants of cancer-related genes in sporadic pheochromocytoma. *International Journal of Genomics* **2018** (2018).
35 Becker, A. P. *et al.* KIAA1549: BRAF gene fusion and FGFR1 hotspot mutations are prognostic factors in pilocytic astrocytomas. *Journal of Neuropathology & Experimental Neurology* **74**, 743-754 (2015).
36 Geijtenbeek, T. B. *et al.* Identification of DC-SIGN, a novel dendritic cell–specific ICAM-3 receptor that supports primary immune responses. *Cell* **100**, 575-585 (2000).
37 Candolfi, M. *et al.* B cells are critical to T-cell—mediated antitumor immunity induced by a combined immune-stimulatory/conditionally cytotoxic therapy for glioblastoma. *Neoplasia* **13**, 947-IN923 (2011).
38 Kunisawa, J. *et al.* Sphingosine 1-phosphate–dependent trafficking of peritoneal B cells requires functional NFκB-inducing kinase in stromal cells. *Blood, The Journal of the American Society of Hematology* **111**, 4646-4652 (2008).
39 Brightbill, H. D. *et al.* Conditional Deletion of NF-κB–inducing kinase (NIK) in adult mice disrupts mature B cell survival and activation. *The Journal of Immunology* **195**, 953-964 (2015).

40 Nathoo, N., Barnett, G. & Golubic, M. The eicosanoid cascade: possible role in gliomas and meningiomas. *Journal of clinical pathology* **57**, 6-13 (2004).
41 Miska, J. & Chandel, N. S. Targeting fatty acid metabolism in glioblastoma. *The Journal of Clinical Investigation* **133** (2023).
42 Cruz, A. L., Barreto, E. d. A., Fazolini, N. P., Viola, J. P. & Bozza, P. T. Lipid droplets: platforms with multiple functions in cancer hallmarks. *Cell death & disease* **11**, 105 (2020).
43 Lopatina, T. *et al.* IL-3 signalling in the tumour microenvironment shapes the immune response via tumour endothelial cell-derived extracellular vesicles. *Pharmacological Research* **179**, 106206 (2022).
44 Zaynagetdinov, R. *et al.* Interleukin-5 facilitates lung metastasis by modulating the immune microenvironment. *Cancer research* **75**, 1624-1634 (2015).
45 Khan, S. *et al.* Intrinsic interferon signaling regulates the cell death and mesenchymal phenotype of glioblastoma stem cells. *Cancers* **13**, 5284 (2021).
46 Ko, A. M.-S., Tu, H.-P. & Ko, Y.-C. Systematic Review of the Role of Alpha-Protein Kinase 1 in Cancer and Cancer-Related Inflammatory Diseases. *Cancers* **14**, 4390 (2022).
47 Li, J. *et al.* The relationship between TRAF6 and tumors. *Cancer Cell International* **20**, 1-12 (2020).
48 Salam, R. *et al.* Cellular senescence in malignant cells promotes tumor progression in mouse and patient Glioblastoma. *Nature Communications* **14**, 441 (2023).
49 Vidyarthi, A. *et al.* Predominance of M2 macrophages in gliomas leads to the suppression of local and systemic immunity. *Cancer Immunology, Immunotherapy* **68**, 1995-2004 (2019).
50 Komi, D. E. A. & Redegeld, F. A. Role of mast cells in shaping the tumor microenvironment. *Clinical reviews in allergy & immunology* **58**, 313-325 (2020).
51 Zhang, X. *et al.* Predicting individual prognosis and grade of patients with glioma based on preoperative eosinophil and neutrophil-to-lymphocyte ratio. *Cancer Management and Research*, 5793-5802 (2020).
52 Ren, F. *et al.* The R132H mutation in IDH 1 promotes the recruitment of NK cells through CX 3 CL 1/CX 3 CR 1 chemotaxis and is correlated with a better prognosis in gliomas. *Immunology and cell biology* **97**, 457-469 (2019).
53 Holl, E. K. *et al.* Examining peripheral and tumor cellular immunome in patients with cancer. *Frontiers in Immunology* **10**, 1767 (2019).
54 Domingues, P. H. *et al.* Immunophenotypic identification and characterization of tumor cells and infiltrating cell populations in meningiomas. *The American journal of pathology* **181**, 1749-1761 (2012).
55 Mauldin, I. S. *et al.* Proliferating CD8+ T cell infiltrates are associated with improved survival in glioblastoma. *Cells* **10**, 3378 (2021).
56 Srinivasa, K., Cross, K. A. & Dahiya, S. BRAF alteration in central and peripheral nervous system tumors. *Frontiers in Oncology* **10**, 574974 (2020).
57 Talloa, D. *et al.* BRAF and MEK targeted therapies in pediatric central nervous system tumors. *Cancers* **14**, 4264 (2022).

58 Braicu, C. *et al.* A comprehensive review on MAPK: a promising therapeutic target in cancer. *Cancers* **11**, 1618 (2019).
59 Villanueva, J., Vultur, A. & Herlyn, M. Resistance to BRAF inhibitors: unraveling mechanisms and future treatment options. *Cancer research* **71**, 7137-7140 (2011).
60 Adelmann, C. H. *et al.* Comparative profiles of BRAF inhibitors: the paradox index as a predictor of clinical toxicity. *Oncotarget* **7**, 30453 (2016).
61 Hu-Lieskovan, S., Robert, L., Moreno, B. H. & Ribas, A. Combining targeted therapy with immunotherapy in BRAF-mutant melanoma: promise and challenges. *Journal of Clinical Oncology* **32**, 2248 (2014).
62 Robinson, M. H. *et al.* Subtype and grade-dependent spatial heterogeneity of T-cell infiltration in pediatric glioma. *Journal for immunotherapy of cancer* **8** (2020).
63 Martin, A. M. *et al.* PD-L1 expression in pediatric low-grade gliomas is independent of BRAF V600E mutational status. *Journal of Neuropathology & Experimental Neurology* **79**, 74-85 (2020).
64 Chan, J. Y. *et al.* Multiomic analysis and immunoprofiling reveal distinct subtypes of human angiosarcoma. *The Journal of clinical investigation* **130**, 5833-5846 (2020).
65 Schroth, W. *et al.* Gene expression signatures of BRCAness and tumor inflammation define subgroups of early-stage hormone receptor–positive breast cancer patients. *Clinical Cancer Research* **26**, 6523-6534 (2020).
66 Wang, Z. *et al.* Classification of pediatric gliomas based on immunological profiling: Implications for immunotherapy strategies. *Molecular Therapy-Oncolytics* **20**, 34-47 (2021).
67 Hsu, J. B.-K. *et al.* Radiomic immunophenotyping of GSEA-assessed immunophenotypes of glioblastoma and its implications for prognosis: a feasibility study. *Cancers* **12**, 3039 (2020).
68 Narang, S. *et al.* Tumor image-derived texture features are associated with CD3 T-cell infiltration status in glioblastoma. *Oncotarget* **8**, 101244 (2017).
69 Kim, A. R. *et al.* Absolute quantification of tumor-infiltrating immune cells in high-grade glioma identifies prognostic and radiomics values. *Cancer Immunology, Immunotherapy* **70**, 1995-2008 (2021).
70 Salvalaggio, A. *et al.* Magnetic Resonance Imaging Correlates of Immune Microenvironment in Glioblastoma. *Frontiers in Oncology* **12**, 823812 (2022).
71 Zhou, Q. *et al.* Correlation of tumor-associated macrophage infiltration in glioblastoma with magnetic resonance imaging characteristics: a retrospective cross-sectional study. *Quantitative Imaging in Medicine and Surgery* **13**, 5958 (2023).
72 Fathi Kazerooni, A. *et al.* The Radiogenomic and Spatiogenomic Landscapes of Glioblastoma, and Their Relationship to Oncogenic Drivers. *bioRxiv*, 2022.2012. 2015.517767 (2022).
73 Ater, J. L. *et al.* Randomized study of two chemotherapy regimens for treatment of low-grade glioma in young children: a report from the Children's Oncology Group. *Journal of clinical oncology* **30**, 2641 (2012).

74 de Blank, P., Bandopadhayay, P., Haas-Kogan, D., Fouladi, M. & Fangusaro, J. Management of pediatric low-grade glioma. *Current opinion in pediatrics* **31**, 21 (2019).
75 Serrallach, B. L. *et al.* Neuroimaging in cerebellar ataxia in childhood: A review. *Journal of Neuroimaging* **32**, 825-851 (2022).
76 Shapiro, J. A. *et al.* OpenPBTA: The Open Pediatric Brain Tumor Atlas. *Cell Genomics* (2023).
77 Li, H. *et al.* The sequence alignment/map format and SAMtools. *bioinformatics* **25**, 2078-2079 (2009).
78 Frankish, A. *et al.* GENCODE 2021. *Nucleic acids research* **49**, D916-D923 (2021).
79 Sturm, G., Finotello, F. & List, M. Immunedeconv: an R package for unified access to computational methods for estimating immune cell fractions from bulk RNA-sequencing data. *Bioinformatics for cancer immunotherapy: methods and protocols*, 223-232 (2020).
80 Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology* **15**, 1-21 (2014).
81 Korotkevich, G. *et al.* Fast gene set enrichment analysis. *BioRxiv*, 060012 (2016).
82 Liberzon, A. *et al.* Molecular signatures database (MSigDB) 3.0. *Bioinformatics* **27**, 1739-1740 (2011).
83 Karczewski, K. J. *et al.* The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434-443 (2020).
84 Taliun, D. *et al.* Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *Nature* **590**, 290-299 (2021).
85 Liu, X., Jian, X. & Boerwinkle, E. dbNSFP: a lightweight database of human nonsynonymous SNPs and their functional predictions. *Human mutation* **32**, 894-899 (2011).
86 Liu, X., Li, C., Mou, C., Dong, Y. & Tu, Y. dbNSFP v4: a comprehensive database of transcript-specific functional predictions and annotations for human nonsynonymous and splice-site SNVs. *Genome medicine* **12**, 1-8 (2020).
87 Landrum, M. J. *et al.* ClinVar: improvements to accessing data. *Nucleic acids research* **48**, D835-D844 (2020).
88 Stenson, P. D. *et al.* The Human Gene Mutation Database: building a comprehensive mutation repository for clinical and molecular genetics, diagnostic testing and personalized genomic medicine. *Human genetics* **133**, 1-9 (2014).
89 Fathi Kazerooni, A. *et al.* Automated tumor segmentation and brain tissue extraction from multiparametric MRI of pediatric brain tumors: A multi-institutional study. *Neuro-Oncology Advances* **5**, vdad027 (2023).
90 Monti, S., Tamayo, P., Mesirov, J. & Golub, T. Consensus clustering: a resampling-based method for class discovery and visualization of gene expression microarray data. *Machine learning* **52**, 91-118 (2003).
91 Si, Y., Liu, P., Li, P. & Brutnell, T. P. Model-based clustering for RNA-seq data. *Bioinformatics* **30**, 197-205 (2014).

92 Cantini, L. *et al.* Benchmarking joint multi-omics dimensionality reduction approaches for the study of cancer. *Nature communications* **12**, 124 (2021).
93 Ester, M., Kriegel, H.-P., Sander, J. & Xu, X. A density-based algorithm for discovering clusters in large spatial databases with noise. *KDD'96 Proc. Second Int. Conf. Knowl. Discov. Data Mining.* **96**, 226–231 (1996).
94 Lê, S., Josse, J. & Husson, F. FactoMineR: an R package for multivariate analysis. *Journal of statistical software* **25**, 1-18 (2008).
95 Wickham, H. & Wickham, H. *Data analysis*. (Springer, 2016).
96 Mayakonda, A., Lin, D.-C., Assenov, Y., Plass, C. & Koeffler, H. P. Maftools: efficient and comprehensive analysis of somatic variants in cancer. *Genome research* **28**, 1747-1756 (2018).
97 Fathi Kazerooni, A. *et al.* "Multiparametric MRI Along with Machine Learning Predicts Prognosis and Treatment Response in Pediatric Low-Grade Glioma". pLGG Radioimmunomics and Clinicoradiomics. DOI: 10.5281/zenodo.13942516 (2024).

6 Acknowledgements

This study was funded by National Institutes of Health (NIH) grants 75N91019D00024, Supplement 3U2CHL156291-03S2, and 75N91019D00024, awarded to A.C.R. The use of the Kids First Data Resource Center (DRC) logo in this publication has been authorized.

7 Author Contributions

A.F.K., A.K., A.C.R., and A.N. conceptualized the study. A.F.K. and A.K., designed and implemented the methodology, supervised data analysis, and wrote the manuscript. K.R., M.C.K., D.G., N.K., A.M.F., A.V., J.B.W., D.H., N.K., H.A., A.M., S.B., M.R.L., S.P., A.S., M.K., Y.Z., contributed to data curation and data generation. K.R., M.C.K., V.K., Y.G., Q.L., X.H., and R.J. contributed to data analysis. A.F.K., A.K., A.N., A.C.R., S.M., J.B.F., M.J.F., C.D., Y.S., contributed to the interpretation of data. A.N., A.C.R., P.B.S. provided resources. All authors revised, commented, and approved the final manuscript. A.F.K. and A.K. contributed equally.

8 Competing Interests Statement

The authors declare no competing interests.

9 Tables

Table 1. Summary of demographics and clinical characteristics of the pLGG patients included in this study (collected through the CBTN data repository).

Variable		**Overall (n = 545)**	**Immunological Clusters (n = 494)**	**Radioimmunomic Signature (n = 150)**	**Clinicoradiomic Risk Stratification (n = 201)**
Age Range (months)		0.16 – 253.9	3.06 - 253.9	4.3 - 253.9	4.3 – 253.9
Sex	Female	250	220	67	97
	Male	295	274	83	104
Tumor Location	Basal Ganglia	3	1	-	2
	Brainstem	34	30	6	10
	Cerebellar	237	229	73	86
	Intraventricular	27	25	3	5
	Lobar (i.e., frontal, parietal, temporal, or occipital lobes)	117	98	31	44
	Multi-regional	28	21	6	14
	Suprasellar	65	58	26	34
	Thalamus	16	14	5	6
	Other	18	18	-	-
WHO 2021 Classification	Diffuse astrocytoma, MYB- or MYBL1-altered	8	7	2	3
	Diffuse glioneuronal tumor with oligodendroglioma-like features and nuclear clusters	2	2	1	1
	Diffuse low-grade glioma, MAPK pathway-altered	342	341	127	128
	Ganglioglioma	28	28	5	5
	Glioneuronal and neuronal tumors	6	3	-	3
	Myxoid glioneuronal tumor	3	3	-	-
	Pediatric-type diffuse low-grade gliomas, NOS	68	44	9	33
	Pilocytic astrocytoma	65	44	4	25
	Pleomorphic xanthoastrocytoma	4	4	-	-
	Subependymal giant cell astrocytoma	19	18	2	3

Table 2. Summary of the performances of clinical-only and clinicoradiomic models on all subjects, and a subgroup analysis on patients who did not undergo gross/near-total resection. The performance metrics include Harrell's concordance index (HAR), Uno's concordance index (UNO), and the Integrated Brier Score (IBS)

	HAR	**UNO**	**IBS**
Model 1: Clinical-Only			
Discovery	0.74 (95% CI = [0.67, 0.80])	0.75 (95% CI = [0.67, 0.82])	0.20 (95% CI = [0.16, 0.22])
Replication	0.76	0.79	0.16
Model 2: Clinicoradiomic			
Discovery	0.71 (95% CI = [0.63, 0.79])	0.72 (95% CI = [0.64, 0.80])	0.24 (95% CI = [0.19, 0.29])
Replication	0.77	0.80	0.16
Subgroup Analysis (Subjects who did not undergo gross/near-total resection)			
Model 1: Clinical-only	0.55	0.57	0.24
Model 2: Clinicoradiomic	0.70	0.70	0.18

10 Figure Legends

Figure 1. Graphical description of the primary analyses in this study. (A) Overview of the imaging, transcriptomic, and multi-modal (combination of imaging and transcriptomic) cohorts. DNA and MRI icons are created by Freepik (DNA: https://www.flaticon.com/free-icons/dna; MRI: https://www.flaticon.com/free-icons/mri) (B) Identification of the immunological clusters within pLGG. (C) Generating a radioimmunomic signature to predict immunological clusters. (D) Developing a clinicoradiomic model that incorporates radiomic features and clinical data to predict progression-free survival in pLGG and thereby, patients' risk of progression, stratify risk, and analyze the biological and immune pathways linked to clinicoradiomic risk. *[Abbreviations: pLGG: pediatric low-grade glioma]*

Figure 2. Immunological clustering. (A) Heatmap indicating the expression levels of immune cells across different patients in the three identified immunological profiles using xCell method (indicated above the heatmap as “cluster assigned”, along with the pLGG molecular subtypes for this cohort of subjects. Color bar represents the xCell enrichment score. (B) Radar plots indicating the upregulated and downregulated set of cell types in each of the immunological clusters. Source data for this figure are provided as a Source Data file. *[Abbreviations: pLGG: pediatric low-grade glioma]*. Source data for this figure are provided as a Source Data file.

Figure 3. Prognostic, clinical, molecular, and immune biomarker associations with immunological clusters. (A) Kaplan-Meier curves representing the probability of PFS

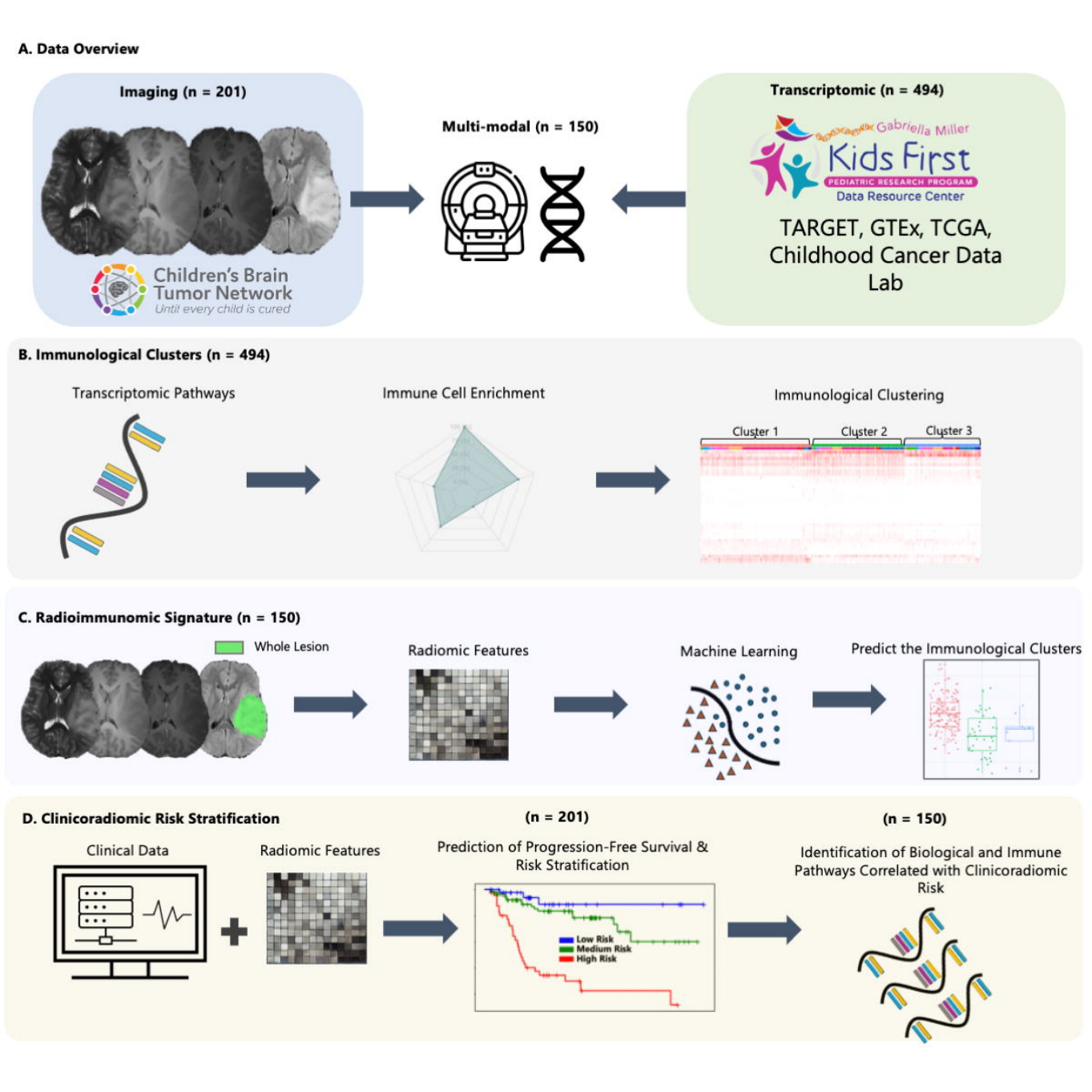
across different immunological clusters, suggesting a statistically significant difference (log-rank test, p = 1.1e-05) between the clusters based on log-rank comparison. The 95% confidence interval of the Kaplan-Meier survival function estimate is provided for each cluster. (B) Forest plot from Cox regression analysis illustrating the effects of immunological cluster assignment, molecular subtypes, age, and sex on PFS, where *N* corresponds to the number of study participants. Error bars represent the 95% confidence interval of the coefficient point estimate from the Cox regression model. Nominal p-values for the coefficients derived from the two-sided Wald test are reported. (C) The heatmap of TIS as a composite of 18 genes. The association with the molecular subtypes and xCell clusters is indicated at the top of the heatmap plot (Cluster 1: n = 188, Cluster 2: n = 163, Cluster 3: n = 137). Color bar represents the z-scores of the TIS scores. (D-E) Violin plots representing the distribution of the values for TIS and TMB, respectively, across the three immunological clusters (Cluster 1: n = 188, Cluster 2: n = 163, Cluster 3: n = 137). The width of each violin represents the density of TMB values for each cluster, highlighting the distribution's shape. The thick black line in the center of each violin represents the median, and the box around it indicates the interquartile range (IQR, from the 25th to 75th percentile). The whiskers are represented with thin black lines extending to the smallest and largest values within 1.5 times the IQR from the lower and upper quartiles. Outliers are shown as individual points outside this range. Statistical comparisons between clusters are displayed above the plot, with the two-sided Kruskal-Wallis test indicating significant differences (p < 2.2e-16 and p = 9.e-06). (F) Correspondence analysis of Pearson residuals derived from Poisson GLM illustrating associations across immune clusters and 2021 WHO CNS tumor entities.

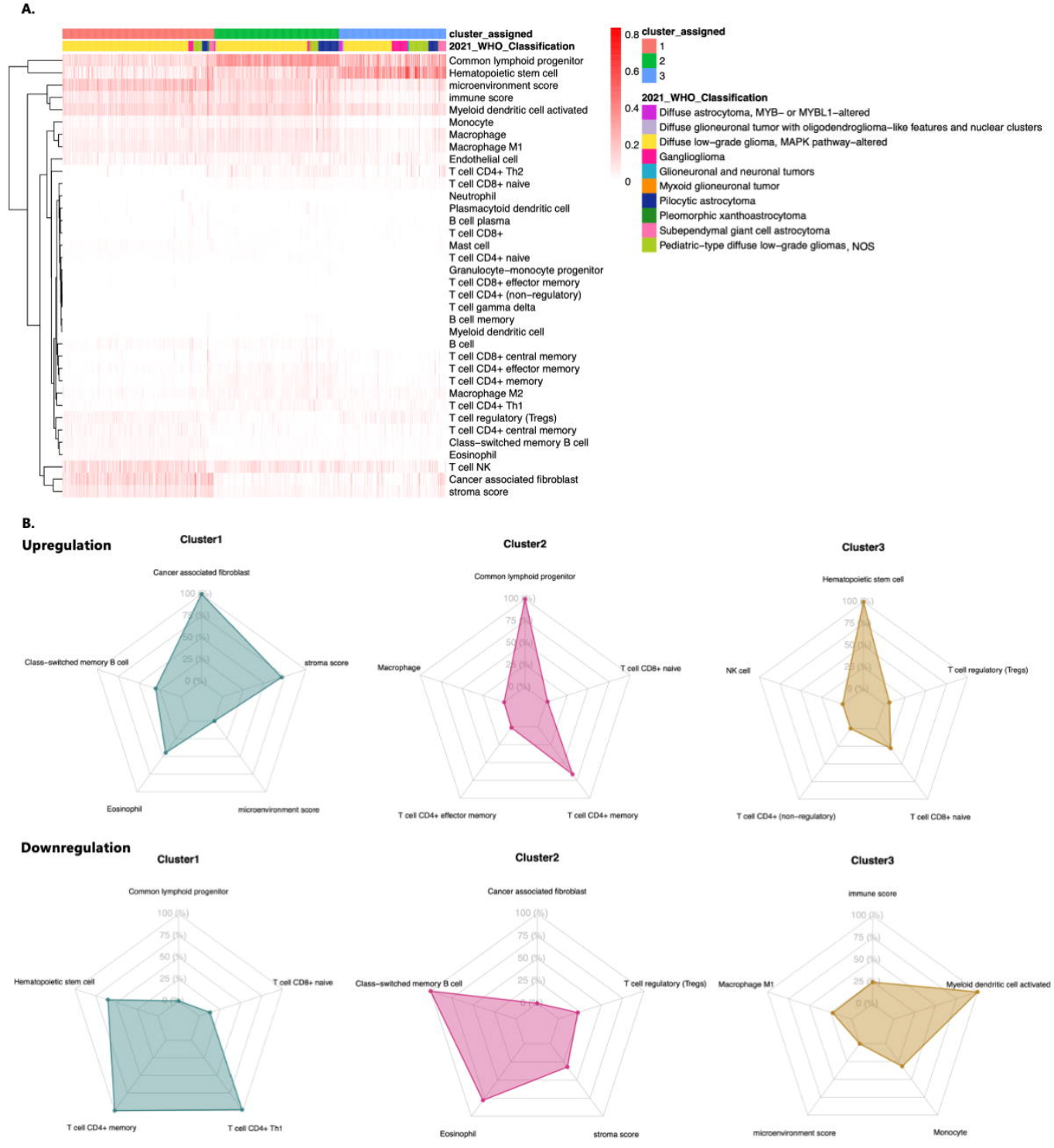
[Abbreviations: PFS: progression-free survival; TIS: tumor inflammation signature; TMB: tumor mutational burden; IQR: interquartile range; GLM: generalized linear model; WHO: World Health Organization; CNS: central nervous system]. Source data for this figure are provided as a Source Data file.

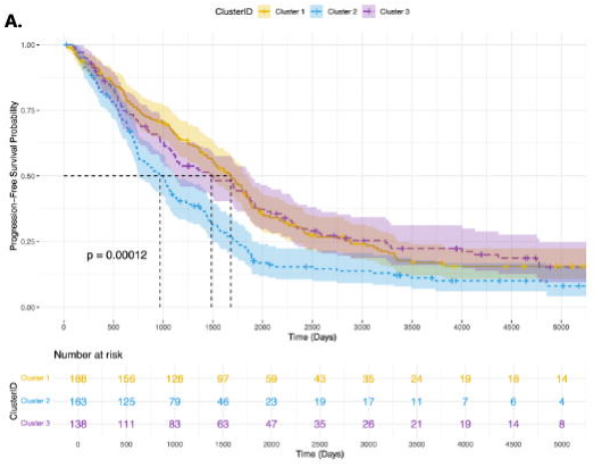
Figure 4. Imaging characteristics associated with immunological clusters. (A) Box-and-Whisker plots indicating radioimmunomic signatures generated for classification of Immune Cluster 2 (n = 47, blue) versus Clusters 1 and 3 (n = 103, red), based on conventional MRI features, as well as conventional MRI + ADC features (Cluster 2, n = 22, blue; Clusters 1 and 3, n = 69, red). The center line represents the median, the box bounds represent the interquartile range (IQR, from the 25th to 75th percentile). The whiskers extend to the smallest and largest values within 1.5 times the IQR from the lower and upper quartiles, with outliers represented as individual points outside this range. (B) Histograms of image intensities within various tumorous subregions for the tumors in immunological clusters 1 (n = 71, green), 2 (n = 47, blue), 3 (n = 32, red). *[Abbreviations: T1-Gd: post-contrast T1-weighted imaging; ADC: apparent diffusion coefficient; CC: cystic component; ED: edema]*. Source data for this figure are provided as a Source Data file.

Figure 5. Clinicoradiomic risk stratification and its association with treatment response and transcriptomic pathways. (A) Kaplan-Meier Curves for the PFS Probability of Clinicoradiomic Risk Scores in the Discovery (n = 160, log-rank test, p = 2.69e-11) and Replication (n = 41, log-rank test, p = 0.0049) cohorts. (B) Box-and-Whisker plots indicating the distribution of risk scores in the patients that received one or no

“systemic” treatments (low-risk, n = 31) (see section 2.3.1 in the manuscript for the definition of the two treatment risk categories), compared to those who received more than one treatment (high-risk, n = 30) over the course of five years (Student’s t-test, p = 0.001). The center line represents the median, the box bounds represent the interquartile range (IQR, from the 25th to 75th percentile). The whiskers extend to the smallest and largest values within 1.5 times the IQR from the lower and upper quartiles, with outliers represented as individual points outside this range. (C) Association plot of standardized residuals for the relationship between the clinicoradiomic and the treatment risk categories (the numbers within each circle in the association plot represent the standardized residuals from the chi-square test of independence). Plots B-C only present the data for the subjects that had progression over a course of five years (n = 61). (D) Sankey plot illustrating the association between the Clinicoradiomic risk groups and the immunological clusters. (E) Bar plot of GLM coefficients derived from an elastic net regression model illustrating pathways associated with the predicted clinicoradiomic risk. *[Abbreviations: PFS: progression-free survival; IQR: interquartile range; GLM: generalized linear model]*. Source data for this figure are provided as a Source Data file.

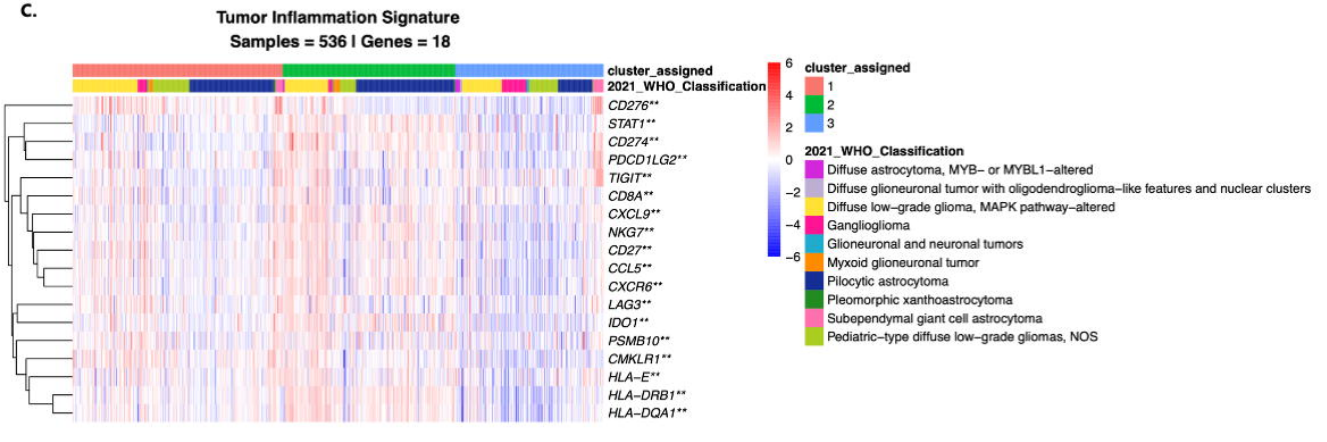

A. Data Overview
Imaging (n = 201)
Children's Brain Tumor Network
Until every child is cured
Multi-modal (n = 150)
Transcriptomic (n = 494)
Gabriella Miller Kids First
PEDIATRIC RESEARCH PROGRAM
Data Resource Center
TARGET, GTEx, TCGA, Childhood Cancer Data Lab
B. Immunological Clusters (n = 494)
Transcriptomic Pathways
Immune Cell Enrichment
Immunological Clustering
Cluster 1
Cluster 2
Cluster 3
C. Radioimmunomic Signature (n = 150)
Whole Lesion
Radiomic Features
Machine Learning
Predict the Immunological Clusters
D. Clinicoradiomic Risk Stratification
(n = 201)
(n = 150)
Clinical Data
Radiomic Features
Prediction of Progression-Free Survival & Risk Stratification
Low Risk
Medium Risk
High Risk
Identification of Biological and Immune Pathways Correlated with Clinicoradiomic Risk

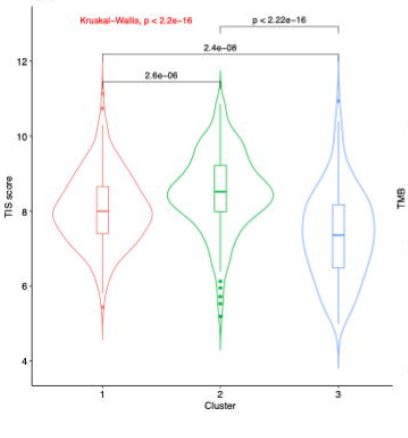

A.
cluster_assigned
2021_WHO_Classification
Common lymphoid progenitor
Hematopoietic stem cell
microenvironment score
immune score
Myeloid dendritic cell activated
Monocyte
Macrophage
Macrophage M1
Endothelial cell
T cell CD4+ Th2
T cell CD8+ naive
Neutrophil
Plasmacytoid dendritic cell
B cell plasma
T cell CD8+
Mast cell
T cell CD4+ naive
Granulocyte−monocyte progenitor
T cell CD8+ effector memory
T cell CD4+ (non−regulatory)
T cell gamma delta
B cell memory
Myeloid dendritic cell
B cell
T cell CD8+ central memory
T cell CD4+ effector memory
T cell CD4+ memory
Macrophage M2
T cell CD4+ Th1
T cell regulatory (Tregs)
T cell CD4+ central memory
Class−switched memory B cell
Eosinophil
T cell NK
Cancer associated fibroblast
stroma score
0.8
0.6
0.4
0.2
0
cluster_assigned
1
2
3
2021_WHO_Classification
Diffuse astrocytoma, MYB− or MYBL1−altered
Diffuse glioneuronal tumor with oligodendroglioma−like features and nuclear clusters
Diffuse low−grade glioma, MAPK pathway−altered
Ganglioglioma
Glioneuronal and neuronal tumors
Myxoid glioneuronal tumor
Pilocytic astrocytoma
Pleomorphic xanthoastrocytoma
Subependymal giant cell astrocytoma
Pediatric−type diffuse low−grade gliomas, NOS
B.
Upregulation
Cluster1
Cancer associated fibroblast
stroma score
microenvironment score
Eosinophil
Class−switched memory B cell
Cluster2
Common lymphoid progenitor
T cell CD8+ naive
T cell CD4+ memory
T cell CD4+ effector memory
Macrophage
Cluster3
Hematopoietic stem cell
T cell regulatory (Tregs)
T cell CD8+ naive
T cell CD4+ (non−regulatory)
NK cell
Downregulation
Cluster1
Common lymphoid progenitor
T cell CD8+ naive
T cell CD4+ Th1
T cell CD4+ memory
Hematopoietic stem cell
Cluster2
Cancer associated fibroblast
T cell regulatory (Tregs)
stroma score
Eosinophil
Class−switched memory B cell
Cluster3
immune score
Myeloid dendritic cell activated
Monocyte
microenvironment score
Macrophage M1
100 (%)
75 (%)
50 (%)
25 (%)
0 (%)

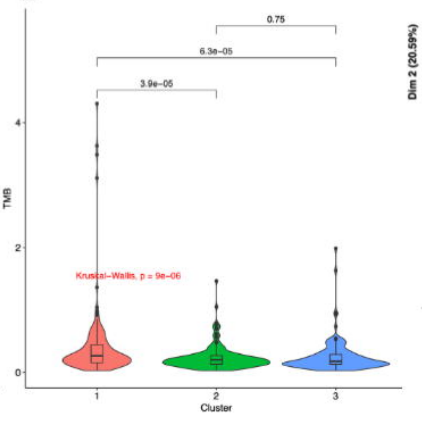


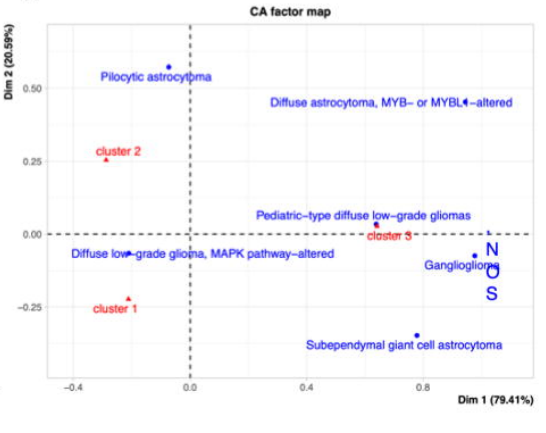
B.

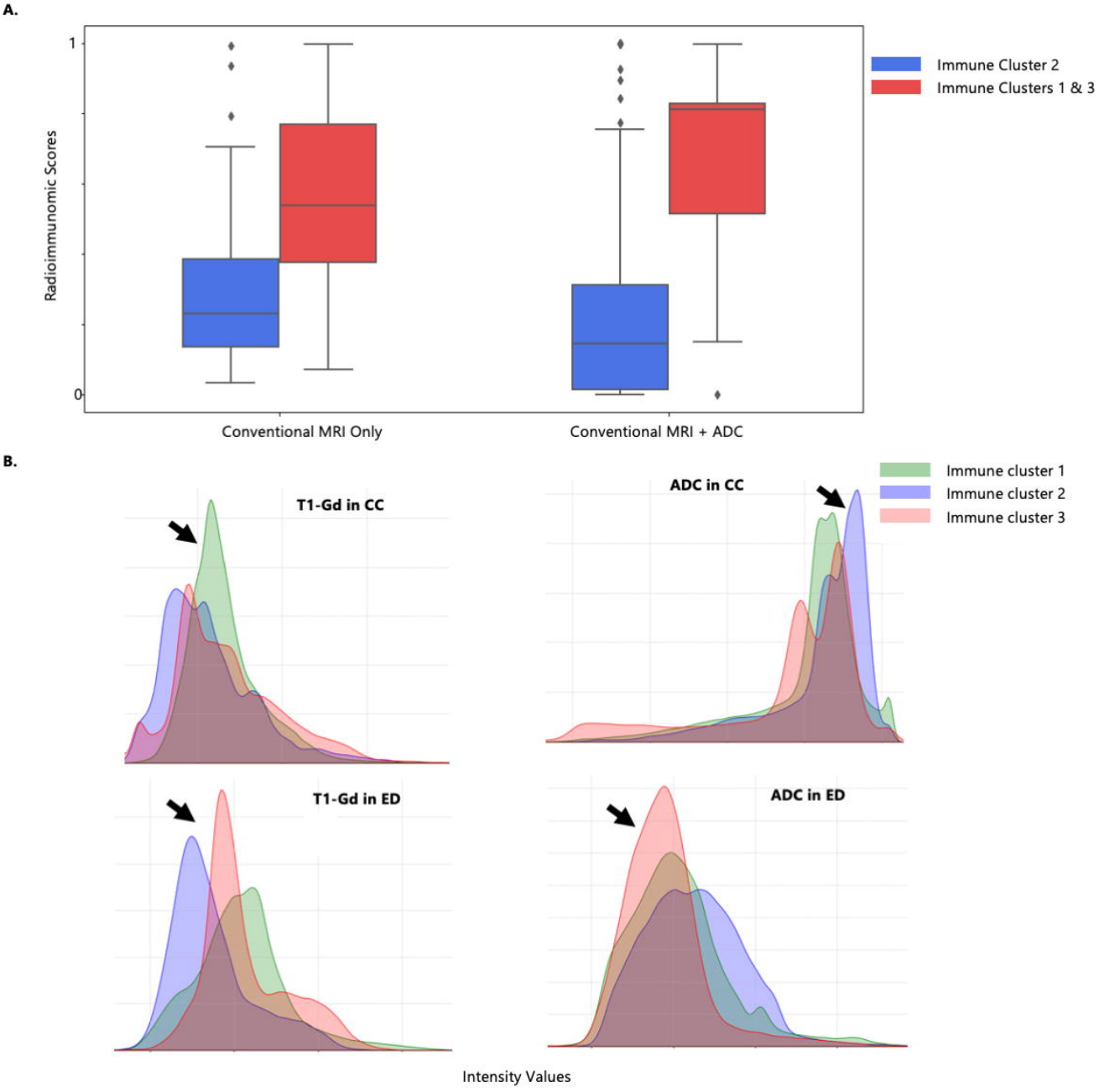
Variable		N	Hazard ratio	p
Age at diagnosis (Days)		488	1.00 (1.00, 1.00)	0.002
Biological Sex	Female	219	Reference	
	Male	269	0.98 (0.79, 1.21)	0.836
Immune Cluster	3	137	Reference	
	1	188	0.95 (0.72, 1.26)	0.721
	2	163	1.50 (1.11, 2.02)	0.008
2021_WHO_Classification	Pediatric-type diffuse low-grade gliomas, NOS	43	Reference	
	Diffuse astrocytoma, MYB- or MYBL1-altered	7	1.47 (0.59, 3.67)	0.404
	Diffuse low-grade glioma, MAPK pathway-altered	339	1.16 (0.77, 1.75)	0.467
	Ganglioglioma	28	1.39 (0.78, 2.47)	0.264
	Glioneuronal and neuronal tumors	3	2.03 (0.47, 8.75)	0.342
	Myxoid glioneuronal tumor	3	3.20 (0.90, 11.33)	0.072
	Pilocytic astrocytoma	44	1.16 (0.69, 1.95)	0.580
	Pleomorphic xanthoastrocytoma	3	2.53 (0.76, 8.48)	0.132
	Subependymal giant cell astrocytoma	18	0.62 (0.25, 1.50)	0.288

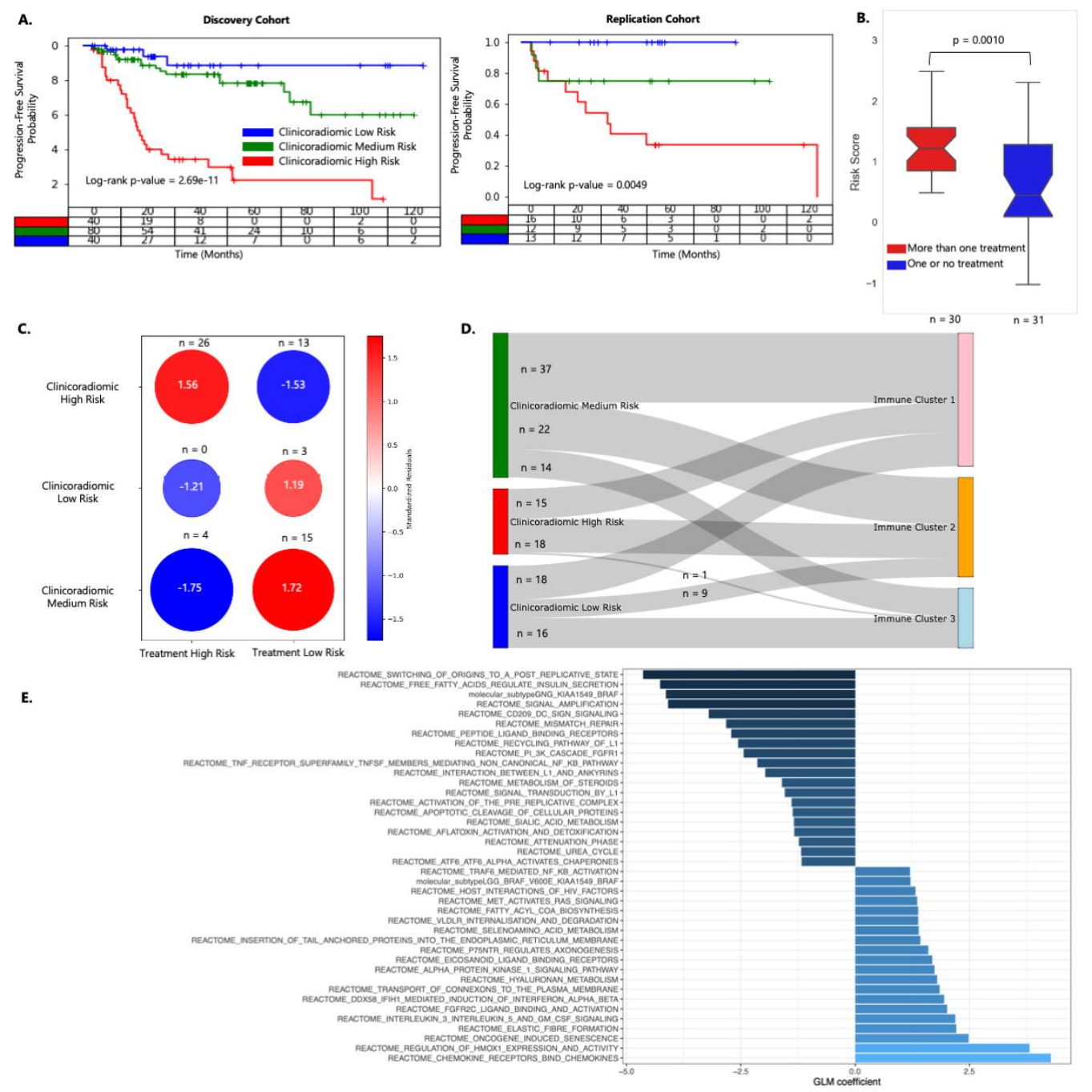









A.
1
0
Radioimmunomic Scores
Conventional MRI Only
Conventional MRI + ADC
Immune Cluster 2
Immune Clusters 1 & 3
B.
T1-Gd in CC
ADC in CC
T1-Gd in ED
ADC in ED
Immune cluster 1
Immune cluster 2
Immune cluster 3
Intensity Values


A.
Discovery Cohort
Progression-Free Survival Probability
Clinicoradiomic Low Risk
Clinicoradiomic Medium Risk
Clinicoradiomic High Risk
Log-rank p-value = 2.69e-11
Time (Months)
Replication Cohort
Log-rank p-value = 0.0049
B.
p = 0.0010
Risk Score
More than one treatment
One or no treatment
n = 30
n = 31
C.
Clinicoradiomic High Risk
Clinicoradiomic Low Risk
Clinicoradiomic Medium Risk
Treatment High Risk
Treatment Low Risk
Standardized Residuals
D.
Clinicoradiomic Medium Risk
Clinicoradiomic High Risk
Clinicoradiomic Low Risk
Immune Cluster 1
Immune Cluster 2
Immune Cluster 3
E.
REACTOME_SWITCHING_OF_ORIGINS_TO_A_POST_REPLICATIVE_STATE
REACTOME_FREE_FATTY_ACIDS_REGULATE_INSULIN_SECRETION
molecular_subtypeGNG_KIAA1549_BRAF
REACTOME_SIGNAL_AMPLIFICATION
REACTOME_CD209_DC_SIGN_SIGNALING
REACTOME_MISMATCH_REPAIR
REACTOME_PEPTIDE_LIGAND_BINDING_RECEPTORS
REACTOME_RECYCLING_PATHWAY_OF_L1
REACTOME_PI_3K_CASCADE_FGFR1
REACTOME_TNF_RECEPTOR_SUPERFAMILY_TNFSF_MEMBERS_MEDIATING_NON_CANONICAL_NF_KB_PATHWAY
REACTOME_INTERACTION_BETWEEN_L1_AND_ANKYRINS
REACTOME_METABOLISM_OF_STEROIDS
REACTOME_SIGNAL_TRANSDUCTION_BY_L1
REACTOME_ACTIVATION_OF_THE_PRE_REPLICATIVE_COMPLEX
REACTOME_APOPTOTIC_CLEAVAGE_OF_CELLULAR_PROTEINS
REACTOME_SIALIC_ACID_METABOLISM
REACTOME_AFLATOXIN_ACTIVATION_AND_DETOXIFICATION
REACTOME_ATTENUATION_PHASE
REACTOME_UREA_CYCLE
REACTOME_ATF6_ATF6_ALPHA_ACTIVATES_CHAPERONES
REACTOME_TRAF6_MEDIATED_NF_KB_ACTIVATION
molecular_subtypeLGG_BRAF_V600E_KIAA1549_BRAF
REACTOME_HOST_INTERACTIONS_OF_HIV_FACTORS
REACTOME_MET_ACTIVATES_RAS_SIGNALING
REACTOME_FATTY_ACYL_COA_BIOSYNTHESIS
REACTOME_VLDLR_INTERNALISATION_AND_DEGRADATION
REACTOME_SELENOAMINO_ACID_METABOLISM
REACTOME_INSERTION_OF_TAIL_ANCHORED_PROTEINS_INTO_THE_ENDOPLASMIC_RETICULUM_MEMBRANE
REACTOME_P75NTR_REGULATES_AXONOGENESIS
REACTOME_EICOSANOID_LIGAND_BINDING_RECEPTORS
REACTOME_ALPHA_PROTEIN_KINASE_1_SIGNALING_PATHWAY
REACTOME_HYALURONAN_METABOLISM
REACTOME_TRANSPORT_OF_CONNEXONS_TO_THE_PLASMA_MEMBRANE
REACTOME_DDX58_IFIH1_MEDIATED_INDUCTION_OF_INTERFERON_ALPHA_BETA
REACTOME_FGFR2C_LIGAND_BINDING_AND_ACTIVATION
REACTOME_INTERLEUKIN_3_INTERLEUKIN_5_AND_GM_CSF_SIGNALING
REACTOME_ELASTIC_FIBRE_FORMATION
REACTOME_ONCOGENE_INDUCED_SENESCENCE
REACTOME_REGULATION_OF_HMOX1_EXPRESSION_AND_ACTIVITY
REACTOME_CHEMOKINE_RECEPTORS_BIND_CHEMOKINES
GLM coefficient