The recent expansion of publicly available childhood cancer datasets offers a tremendous opportunity to deepen our understanding of tumor biology and improve outcomes for children with rare, understudied tumor types such as pediatric brain cancers. Our project ideas aim to enhance existing data and web-based platforms that support exploration of integrated pediatric brain cancer datasets, including those from the CCDI Molecular Characterization Initiative and the Children’s Brain Tumor Network. We welcome the opportunity to collaborate on any or all aspects of the ideas outlined in this abstract.

We envision developing user-friendly interactive data visualization tools such as oncoprints, survival plots, and mutation heatmaps that enable researchers without coding or bioinformatics expertise to dynamically create and analyze cohorts. These cohorts will encompass clinical, genomic, pathology, treatment, and follow-up data. By lowering barriers to complex data access and enabling intuitive exploration, the proposed tool will empower clinicians, pathologists, and cancer researchers to identify patterns and generate hypotheses, particularly in the context of rare or newly defined pediatric brain tumor types.

A critical component of central nervous system (CNS) tumor classification in pediatric and young adult populations is DNA methylation profiling. However, existing methylation datasets within CCDI have been processed using different classifier versions and reference cohorts, reflecting the evolving nature of the field. To address this inconsistency, we propose re-processing all available raw methylation data files (.idat files) using one or more current classifiers (e.g., NCI Bethesda and/or IGM v1.0). In addition to standardizing the available data, employing multiple classifiers enhances reproducibility and comparability, facilitates standardization across future classifier versions, and increases diagnostic confidence particularly in borderline or ambiguous cases. Additionally, newer classifiers may include refined or very rare subtypes that earlier versions did not capture in their reference set. This harmonized classification output could be made available through dbGaP to support cross-study comparisons and maximize data utility.

In addition to data standardization, visualization, and cohort building, we propose leveraging whole slide pathology images to develop algorithms for histopathologic feature extraction and machine learning applications. These tools could significantly improve diagnostic precision—especially critical given the nationwide shortage of pediatric neuropathologists. Researchers will be able to define and compare subgroups within larger cohorts to explore morphological differences and identify potential diagnostic or prognostic markers.

This integrated approach of combining clinical-genomic-methylation data with advanced image analysis has the potential to accelerate discovery in rare brain tumor types, support accurate diagnoses in resource-limited settings, and inform clinical trial design for treatment stratification and precision medicine in pediatric neuro-oncology.

Our multidisciplinary team brings expertise in clinical neuro-oncology, neuropathology, advanced clinical genomic testing, statistics, and bioinformatics. To achieve the goals of this proposal, we hope to collaborate with experts in data visualization and machine learning to develop accessible, interactive tools for the broader pediatric brain tumor research community.