Genetic ancestry superpopulations show distinct prevalence and outcomes across pediatric central nervous system tumors

Ryan J. Corbett[1,2,3], Cricket C. Gullickson[4], Zhuangzhuang Geng[1,2], Miguel A. Brown[1,2], Bo Zhang[1,2], Chuwei Zhong[1,2], Nicholas Van Kuren[1,2], Antonia Chroni[1,2,+], Christopher Blackden[1,2], Ammar S. Naqvi[1,2], Alexa Plisiewicz[1,2,++], Sean McHugh[1,2], Emmett Drake[1,2,#], Kaitlin Lehmann[1,2], Tom B. Davidson[5], Michael Prados[6], Phillip B. Storm[1,2], Adam C. Resnick[1,2], Angela J. Waanders[7,8], Sebastian M. Waszak[6,9], Sabine Mueller[6], Jo Lynne Rokita[1,2,10,*,^], Cassie Kline[3,4,11,*,^]

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

[2]Division of Neurosurgery, Children's Hospital of Philadelphia, Philadelphia, PA, USA

[3]Division of Oncology, Children's Hospital of Philadelphia, Philadelphia, PA, USA

[4]Department of Pediatrics, Children's Hospital of Philadelphia, Philadelphia, PA, USA

[5]Keck School of Medicine of University of Southern California, Children's Hospital of Los Angeles, Los Angeles, CA, USA

[6]Department of Neurology, University of California, San Francisco, San Francisco, California, USA

[7]Division of Hematology, Oncology, Neuro-Oncology, and Stem Cell Transplant, Ann & Robert H Lurie Children's Hospital of Chicago, USA

[8]Department of Pediatrics, Northwestern University Feinberg School of Medicine

[9]Laboratory of Computational Neuro-Oncology, Swiss Institute for Experimental Cancer Research, School of Life Sciences, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland

[10]Department of Biomedical and Health Informatics, Children's Hospital of Philadelphia, Philadelphia, PA, USA

[11]Department of Pediatrics, University of Pennsylvania, Philadelphia, PA, USA

[+]Present affiliation: Department of Development Neurobiology, St. Jude Children's Research Hospital, Memphis, TN, USA

[#]Present affiliation: Sidney Kimmel Cancer Center, Thomas Jefferson University Hospital, Philadelphia, PA, USA

[++]Present affiliation: Division of Nephrology, University of Washington, Seattle, WA, USA

*Senior author equal contribution

[^]Correspondence:

Cassie Kline, M.D., M.A.S., klinec@chop.edu

Jo Lynne Rokita, Ph.D., rokita@chop.edu

Abstract

Background: Central nervous system (CNS) tumors lead to cancer-related mortality in children. Genetic ancestry-associated cancer prevalence and outcomes have been studied, but is limited.

Methods: We performed genetic ancestry prediction in 1,484 pediatric patients with paired normal and tumor whole genome sequencing from the Open Pediatric Cancer (OpenPedCan) project to evaluate the influence of reported race and ethnicity and ancestry-based genetic superpopulations on tumor histology, molecular subtype, survival, and treatment.

Results: Predicted superpopulations included African (AFR, N=155), Admixed American (AMR, N=224), East Asian (EAS, N=67), European (EUR, N=995), and South Asian (SAS, N=43). Reported race and ethnicity and ancestry-based genetic superpopulations were non-randomly associated. Patients with an atypical teratoid rhabdoid tumor (ATRT), *MYC* subtype or meningioma were enriched for AFR ancestry. Patients of AMR ancestry with *KIAA1549::BRAF* fusion-positive low-grade glioma (LGG) had tumors enriched for rare fusion breakpoints, lesser extent of surgical resection, and worse event-free survival (EFS). Non-EUR and AMR patients with germ cell tumors or SHH-activated medulloblastoma exhibited worse EFS relative to EUR patients, and patients of AFR ancestry with LGG or ependymoma had worse overall survival compared to EUR patients. We observed higher frequency of clinical trial enrollment among AMR patients across tumor histologies, but increased utilization of photon versus proton radiation relative to other superpopulations.

Conclusions: Genetic ancestry-associated disparities exist across pediatric CNS tumor histological and molecular subtypes. Further investigation into genetic and socioeconomic factors contributing to these observed inequities is needed.

Key Words: Pediatric neuro-oncology, Health equity, Clinical Outcomes, Genetic Ancestry

Key Points: Distinct associations of genetic ancestry-based superpopulations exist within pediatric CNS tumor histologic and molecular subtypes and correlate with survival outcomes and treatment.

Importance of the Study: This work provides critical insight on the impact of reported race and ethnicity and genetic-based ancestry superpopulations on nearly 1,500 pediatric patients with CNS tumors who had matched normal and tumor sequencing performed. We identify novel associations between ancestry superpopulations and tumor histology, molecular subtypes, and treatments received. Here, we begin to inform on the contributions of social constructs of race and ethnicity and tumor characteristics that are enriched among genetic-based ancestry superpopulations on clinical outcomes of pediatric patients with CNS tumors. Our findings indicate that potential social and genetic risk stratifications exist for pediatric CNS tumors and warrant further investigation to ensure equitable clinical outcomes for all patients.

Introduction

Primary central nervous system (CNS) tumors are the leading cause of cancer-related mortality in children[1,2]. Previous work has shown that overall and tumor-specific survival outcomes and incidence vary according to race and ethnicity in children with primary CNS tumors[1,2]. For example, while White children have a higher incidence of CNS tumors in general, they have a lower incidence of malignant CNS tumors compared to other races[1,3]. Furthermore, Black and Hispanic patients have higher rates of mortality compared to White patients and Hispanic children are more likely to present with advanced diseases[1,2,4]. Prior studies have attempted to characterize the contribution of sociodemographic factors, such socioeconomic status, on survival outcomes and treatment strategies while accounting for key variables like extent of disease, type of treatment, and age at diagnosis. Even in consideration of these confounding variables, findings demonstrate continued differences in survival and treatment according to patient race and ethnicity[3,5,6]. This suggests that unmeasured social determinants of health (SDoH) or inherent genetic variation in cancer risk may be playing a role[3]. What remains less well understood is the individual contribution of these disparate outcomes as it relates to pediatric CNS tumor type, molecular subtype, and clinical characteristics.

It is essential to recognize that race and ethnicity are social and cultural constructs distinct from genetic ancestry, which can be estimated using genetic markers that capture ancestral population migration patterns and admixture events[7]. The use of predicted genetic ancestry in cancer studies has revealed numerous ancestry-based correlates to cancer incidence and outcome. Work using data from The Cancer Genome Atlas (TCGA) showed increased frequency of *TP53* mutations in patients of African ancestry with cancer types demonstrating high chromosomal instability, as well as decreased frequency of *VHL* and *PBRM1* mutations in renal cancer patients of African ancestry[8,9]. In a study of pediatric acute lymphoblastic leukemia (ALL), East Asian ancestry was negatively associated with incidence of *BCR-ABL1*-like and T-cell ALL, while increased proportion of African and Native American genetic ancestry was associated with

worse overall and event-free survival (OS and EFS, respectively)[10]. While concordance between reported race and ethnicity and predicted genetic ancestry can vary by group, prior work has reported a significant non-random association between these categories[11]. Thus, the use of genetic ancestry to assess demographic inequities in cancer outcomes may be particularly useful in cohorts for which electronic medical records (EMRs) are incomplete or inaccurate[7].

In the current study, we aim to go beyond previous work investigating genomic correlates of cancer risk in isolated silos and utilize a large cohort of patients with broad histologies of primary pediatric CNS tumors to explore the potential contributions of genomics and social health risk categories. We specifically utilize reported race and ethnicity and perform genetic ancestry prediction to evaluate associations with incidence and clinical outcomes across a diverse group of pediatric CNS tumor diagnoses. This work provides an essential framework to better characterize the contributions of genetic and sociodemographic factors to cancer outcomes in patients with pediatric CNS tumors, and ideally augment our understanding of pediatric CNS tumor risk stratification through a lens of health equity.

Materials & Methods

Pediatric CNS tumor patient cohort

The pediatric CNS tumor patient cohort used in this study was derived from the Open Pediatric Cancer (OpenPedCan) project[12,13], an open analysis effort at the Children's Hospital of Philadelphia that performs pediatric cancer data harmonization and shares results from downstream analyses. Patients with matched tumor and normal whole genome sequencing (WGS) from OpenPedCan at the time of data release v12 were included for this study. These included patients from the Children's Brain Tumor Network (CBTN, https://cbtn.org/) and the Pediatric Neuro-oncology Consortium (PNOC, https://pnoc.us/). Patients from countries outside the United States were excluded due to consent procedures commonly necessary to ascertain race and ethnicity data as per local regulatory requirements. All specimen- and participant-

associated IDs were randomly generated and de-identified for each patient. We pulled demographics and clinical data such as reported race and ethnicity, initial tumor diagnosis, tumor histology, molecular subtype, survival outcomes (EFS and OS), and selected treatment data from OpenPedCan data release v14. A detailed overview of molecular subtyping methods can be found in Shapiro et al. 2023[12] with updates to high-grade glioma (HGG) and atypical teratoid rhabdoid tumor (ATRT) described in OpenPedCan[13]. Collected treatment information through the IRB-approved CBTN project included extent of tumor resection, utilization of upfront proton versus photon therapy for those who received radiation therapy, and clinical trial enrollment at time of diagnosis. Patient data from the PNOC cohort, such as demographics, diagnostic data, molecular sequencing, and survival, was collected within confines of clinical trial data collection and shared as per IRB-approved consent. Reported race and ethnicity were utilized as entered by local site investigators and research teams.

Predicted ancestry

We used the somalier suite of tools (v.0.2.15) to predict ancestry superpopulations from non-tumor WGS data[14]. First, `somalier-extract` was applied to alignment CRAM files to obtain variant calls at known polymorphic sites for each sample. We then used `somalier-ancestry` with default parameters to estimate ancestry superpopulations in patients using genetic markers from reference individuals of known ancestry from the 1000 Genomes Project[15]. Briefly, dimensionality reduction was performed on query and reference genotype data to estimate five principal components (PCs), and the resulting PC values were used to estimate the proportion of genetic ancestry assigned to each of five ancestry superpopulations: Sub-Saharan African (AFR), Admixed American (AMR), East Asian (EAS), European (EUR), and South Asian (SAS). The ancestry group with the highest estimated percentage in each sample was assigned as the predicted primary genetic ancestry superpopulation.

BRAF fusion breakpoint analyses for low-grade glioma (LGG)

All high-confidence, in-frame *KIAA1549*::*BRAF* STAR-fusion and/or Arriba fusion calls in pediatric low-grade glioma (LGG) tumors were annotated with exon number in canonical transcripts (NM_001164665 for *KIAA1549* and NM_004333 for *BRAF*) using *biomaRt* and *GenomicRanges* R packages[16-18]. Common breakpoints included those involving exons 15:09 (exon 15 in *KIAA1549* and exon 9 in *BRAF*), 16:09, 16:11, and 18:10[19,20]. All other breakpoint combinations were classified as rare or novel.

Statistical analyses

The concordance between reported race and ethnicity and genetic ancestry superpopulation was assessed using Fisher's exact tests. Enrichment of genetic ancestries within race and ethnicity groups was calculated using hypergeometric tests. We integrated CNS tumor histology and molecular subtype data from matched tumor samples with survival data to determine whether genetic ancestry superpopulations were associated with cancer type, subtype, and/or survival. Non-random distribution of predicted ancestry superpopulations within histologies and molecular subtypes was assessed using Fisher's exact tests and enrichment was calculated using hypergeometric tests. Statistical significance was defined as $p<0.05$.

We performed Kaplan-Meier analysis of OS and EFS within histologies and molecular subtypes to compare outcomes of patients of different predicted ancestry superpopulations. We generated Cox proportional-hazards regression models to identify additional variables that were predictive of survival using the following covariates: age at diagnosis (all models), molecular subtype (Atypical Teratoid Rhabdoid Tumor [ATRT], low- and high-grade gliomas [LGG and HGG, respectively], ependymoma [EPN], medulloblastoma [MB], mixed glioneuronal tumors), and extent of tumor resection (LGG only, to accommodate analysis of specific breakpoint differences on outcome and known impact of degree of resection on outcome in this cohort). Analyses of deviance were performed on Cox proportional-hazards models to assess overall effect of genetic

ancestry superpopulation on survival. All survival analyses were performed using the *survival* R package[21], and Kaplan-Meier survival curves were generated using the *survminer* R package[22].

Results

Ancestry superpopulation prediction summary

We predicted genetic ancestry for 1,484 pediatric CNS tumor patients enrolled in the CBTN and/or PNOC. Patients were classified into ancestry superpopulations as follows: N=155 AFR, N=224 AMR, N=67 EAS, N=995 EUR, and N=43 SAS (**Figure 1A, Table S1**). Most patients (1306/1484; 88%) exhibited predicted ancestry superpopulation probabilities greater than 90% (**Figure S1A**). The remaining 178 patients were either predicted to be of AMR superpopulation with a significant secondary superpopulation probability or were predicted to be of non-AMR superpopulation with a significant secondary AMR superpopulation probability (**Figure S1B**).

There was a significant non-random association between reported race and ethnicity and predicted genetic ancestry superpopulation ($p<0.001$; **Figure 1B, Table S2**). Patients of each predicted ancestry superpopulation were significantly enriched for distinct reported race groups (**Figure 1C**). This included AFR patients for Black/African American race, AMR patients for American Indian and Alaska Native (AI/AN) and Native Hawaiian and Pacific Islander (NHPI) races, EAS patients for Asian and NHPI races, EUR patients for White race, and SAS patients for AI/AN and Asian races. AMR patients were significantly enriched among patients of reported Hispanic/Latino ethnicity, while AFR, EUR, and SAS patients were enriched among patients of reported non-Hispanic/Latino ethnicity (**Figure 1D**). We predicted genetic ancestry superpopulations in 343 individuals (23% of all study participants) for which reported race or ethnicity data were not available, and these patients were disproportionately assigned to AMR and SAS superpopulations (AMR: N=149, OR=3.1, p=2.7e-58; SAS: N=22, OR=2.4, p=3.3e-06). All patients with reported race and ethnicity and/or predicted ancestry were included in final analyses.

Genetic ancestry superpopulations are enriched for distinct pediatric CNS tumor histologies

We identified a significant non-random association between genetic ancestry superpopulation and CNS tumor histology (p=1.0e-04; **Figure 2A, Table S2**). Patients with ATRT and meningiomas were significantly enriched within the AFR superpopulation (N=11, OR=2.2, FDR=0.02 and N=10, OR=2.4, FDR=0.02, respectively). We observed a significant enrichment of patients with DIPG or DMG within the AMR superpopulation (N=27, OR=1.7, FDR=0.02). To determine if this enrichment was due to inclusion of patients enrolled in PNOC trials—of which the majority (30/35, 85.7%) had DIPG or DMG—we assessed superpopulation distribution among these patients. The AMR superpopulation was significantly enriched among PNOC patients (N=14, OR=3.9, p=2.6E-04). Furthermore, re-analysis of tumor histology enrichment among superpopulations when excluding PNOC patients resulted in loss of AMR superpopulation enrichment in patients with DIPG or DMG (N=14, OR=1.3, FDR=0.5), indicating that the initial observed enrichment was due to sample bias. Patients with germ cell tumors were significantly enriched among the EAS superpopulation (N=4, OR=3.8, FDR=0.01), with three patients diagnosed with teratomas and one with a mixed germ cell tumor. Patients with LGG and oligodendrogliomas were significantly enriched within the EUR superpopulation (N=275, OR=1.1, FDR=0.046 and N=52, OR=1.2, FDR=0.03, respectively). Lastly, patients with schwannomas were significantly enriched among SAS superpopulation patients (N=4, OR=5.5, FDR=0.01).

Novel molecular findings can be found across genetic ancestry superpopulations

We observed significant enrichment of genetic ancestry superpopulations in molecular subtypes of specific CNS tumor histologies. We found that patients with *BRAF* V600E mutant as well as non-*BRAF* altered LGG ("Other alteration") tumors were enriched among the EUR superpopulation (N=34, OR=1.1, p=0.04 and N=68, OR=1.1, p=0.01, respectively; **Figure 2B**). Patients with EPN *ZFTA* fusion-positive tumors were disproportionately represented in the EAS

superpopulation relative to other superpopulations (N=3, OR=2.1, p=0.02; **Figure 2C**). A significant enrichment of ATRT-MYC subtype tumors was found within the AFR superpopulation (N=5, OR=1.7, p=0.02) and, among patients with glioneuronal tumors, there was a significant enrichment of wildtype (WT) and non-*BRAF* altered subtypes in patients from EUR and SAS superpopulations, respectively (N=15, OR=1.2, p=0.048 and N=5, OR=1.6, p=0.02, respectively; **Figure S2A-D**).

A recent report from the PNOC001 trial identified two different classes of in-frame *KIAA1549*::*BRAF* fusion breakpoints in LGG tumors denoted as "common" (16:09, 15:09, 16:11, and 18:10) or "rare" (any other novel combination), with rare breakpoints associated with supratentorial midline pilocytic astrocytoma and poor clinical outcomes[20]. Since LGGs made up the largest histology of this cohort and we had a large number of patients with *KIAA1549*::*BRAF* fusions (N = 172/383, 44.9%), we sought to determine whether ancestral associations with the breakpoint type exist. All LGG tumors with *KIAA1549::BRAF* fusions were confirmed to have a pilocytic astrocytoma (n=162), pilomyxoid astrocytoma (N=8), or fibrillary astrocytoma (N=2) diagnosis. We annotated *KIAA1549::BRAF* fusion breakpoints as "common" or "rare" to determine if distribution of breakpoint types differed across genetic ancestry superpopulations. The majority of tumors (161/172, 93.6%) harbored one of the four "common" breakpoints, ranging from 87.5% (21/24) of patients from the AMR superpopulation to 100% of patients from EAS and SAS superpopulations. Prevalence of the 15:09 breakpoint was not evenly distributed across predicted ancestries (p=0.04), with patients from EAS and SAS superpopulations being significantly enriched for 15:09 breakpoints relative to other superpopulations (EAS: N=2, OR=3.7, p=5.4e-03; SAS: N=3, OR=3.3, p=3.9e-03; **Figure 3A**). Patients from the AMR superpopulation were significantly enriched for rare/novel *BRAF* fusion breakpoints (N=3, OR=2.0, p=0.049).

We also observed breakpoint type-specific distribution of LGG tumors in CNS regions that correlated with extent of tumor resection (**Figure 3B-C**). Tumors with *KIAA1549::BRAF* 16:09 breakpoints were significantly enriched in the posterior fossa (N=84, OR=1.1, p=4.3e-04), and

these were significantly more likely to undergo total resection than other breakpoint types (N=78, OR=1.1, p=4.3e-04). Conversely, tumors with the *KIAA1549::BRAF* 15:09 breakpoint were significantly enriched in suprasellar and spinal regions (N=4, OR=2.3, p=0.02 and N=4, OR=6.2, p=5.3e-04), and those with rare *KIAA1549::BRAF* breakpoints were significantly enriched in suprasellar and optic pathway regions (N=4, OR=6.3, p=7.0e-05 and N=2, OR=8.7, p=5.3e-04). Both 15:09 and rare *KIAA1549::BRAF* breakpoint tumors were significantly more likely to undergo partial resection relative to tumors with other breakpoint types (15:09: N=13, OR=1.4, p=0.03; rare: N=8, OR=2.4, p=3.4e-04). Lastly, 15:09 and rare *KIAA1549::BRAF* breakpoint tumors were significantly more likely to be diagnosed as pilomyxoid astrocytomas (PMAs) versus pilocytic astrocytomas (PAs; 15:09: N=4, OR=4, p=0.01; rare: N=3, OR=5.2, p=1.0E-03), whereas other common breakpoint tumors were more likely to be diagnosed as PAs (16:09: N=102, OR=1.1, p=2.6E-05; **Figure S3**)

We identified eleven patients with seven distinct rare *KIAA1549::BRAF* breakpoints, two of which were recurrent and unique to a single genetic ancestry superpopulation: *KIAA1549::BRAF* 13:09 breakpoints were identified in tumors from two patients from the AMR superpopulation, and *KIAA1549::BRAF* 15:11 breakpoints were identified in tumors from four patients from the EUR superpopulation (**Figure 3D**). To determine if breakpoint type was associated with survival differences, we generated Kaplan-Meier and Cox proportional hazards models for EFS using breakpoint type as a predictor. Significantly worse EFS persisted in patients with rare *KIAA1549::BRAF* breakpoints relative to those with *KIAA1549::BRAF* 16:11 breakpoints, even when accounting for covariates of degree of tumor resection (1.9 versus 2.9 years, HR=5.3, p=0.03, **Figure 3E-F**).

Patients from non-European genetic ancestry superpopulations have significantly worse event-free and overall survival in a subset of pediatric CNS tumors

We calculated median EFS and OS for each CNS tumor histology and molecular subtype with sufficient sample size (N ≥ 20) across predicted ancestry superpopulations (**Table S3**). Genetic ancestry superpopulation significantly correlated with EFS in patients with germ cell tumors (χ=12.4, p=0.01) and plexiform neurofibromas (χ=7.6, p=0.02). Ancestry superpopulation correlated with OS in patients with mesenchymal tumors (χ=14, p=0.01) and LGGs (χ=14.3, p=0.01). This was most notable in for patients with non-*BRAF* altered LGG (χ=13.3, p<0.01).

Next, we sought to assess pairwise differences in patient survival for genetic ancestry superpopulations within histologies and molecular subtypes. Among patients with germ cell tumors, those from the non-EUR ancestry superpopulations exhibited significantly worse EFS relative to those from the EUR superpopulation (median EFS 0.5 vs. 2.1 years, HR=12, p=0.01; **Figure 4A-B**). This persisted even when accounting for covariates of germ cell tumor histology. Furthermore, while genetic ancestry was not a significant predictor of EFS among all patients with MB (χ=1.3, p=0.87), patients with MB-SHH subtype tumors from the AMR superpopulation exhibited significantly worse EFS relative to those patients from the EUR superpopulation (median EFS 1.0 vs. 2.9 years, HR=4.1, p=0.03; **Figure 4C-D**). We did not find evidence of non-random distribution of the four MB SHH subtypes (1, 2, 3, 4) across superpopulations (p=0.4), indicating that subtype is not responsible for ancestry-associated survival differences.

Among patients with LGG, individuals from the AFR superpopulation exhibited significantly worse OS relative to those from the EUR superpopulation (median OS 2.9 vs. 4.2 years, HR=13, p=0.01; **Figure 4E-F**). While we also observed a significantly higher hazard ratio of OS in patients from the SAS vs. EUR superpopulation (HR=35, p=0.03), this was driven by a single event among the SAS group, and therefore may not replicate in a larger cohort. When only considering non-*BRAF* altered LGG tumors, the same trend of higher hazard ratio of OS was observed in individuals from the AFR superpopulation relative to the EUR superpopulation (HR=49, p=0.01; **Figure S4A&C**). This ancestry-associated difference remained significant even when

accommodating for extent of tumor resection. Non-*BRAF* altered tumors were primarily *NF1* or *MAPK*-altered, but we did not find evidence of significant enrichment for any ancestry superpopulation when we investigated more specific subtypes, likely due to low numbers per histology and subtype (**Figure S5**). For patients with EPN tumors, we observed significantly worse OS in individuals from the AFR compared to EUR superpopulation (median OS 2.6 vs. 4.7 years, HR=5.5, p=0.02; **Figure 4G-H**), particularly when only considering *ZFTA* fusion-positive EPN (median OS 2.4 vs. 5.1 years, HR=30, p=0.01; **Figure S4B&D**).

Genetic ancestry superpopulations exhibit distinct treatment patterns

To explore therapeutic approaches employed across genetic ancestry superpopulations, we investigated degree of surgical intervention, clinical trial enrollment, and therapeutic regimens cohort-wide and within tumor histologies. Among patients with LGG, those from AMR and EAS superpopulations were significantly more likely to receive only biopsies over other surgical interventions when compared to other superpopulations (AMR: N=5, OR=1.8, p=0.04; EAS: N=2, OR=2.8, p=0.03; **Figure 2D**). EAS superpopulation patients exhibited higher incidences of LGG in suprasellar and optic pathway regions (N=2, OR=3.8, p=0.01 and N=1, OR=4.0, p=0.02) and AMR superpopulation patient LGGs were enriched in regions of mixed anatomical location (N=11, OR=1.7, p=0.02), which likely contributes to observed differences in surgical resection by superpopulation (**Figure S3E**).

Overall frequency of upfront clinical trial enrollment was significantly associated with genetic ancestry (p=0.001; **Table 1**), with patients from AMR and EUR superpopulations exhibiting significantly higher and lower enrollment rates, respectively, relative to other superpopulations (AMR: OR=2.1, p=1.7e-04; EUR: OR=0.68, p=0.01; **Figure 5A**). The overall higher rate of AMR superpopulation upfront enrollment was driven in part by patients from DIPG or DMG and EPN tumor cohorts, for which increased rates of AMR superpopulation enrollment were also observed (OR=4.4, p=1.6e-03 and OR=5.3, p=0.01, respectively). Genetic ancestry

was also significantly associated with rate of upfront enrollment among patients with mesenchymal tumors (p=0.048), with patients from the SAS superpopulation having significantly higher rates of enrollment than those of other superpopulations (OR=21.9, p=0.03). And, among patients with HGG, those from the AFR superpopulation were enrolled at significantly higher rates than those from other superpopulations (OR=8.9, p=0.02).

We further assessed whether frequency of proton versus photon radiation therapy differed across genetic ancestry superpopulations. There was a trend toward genetic ancestry-dependent frequency of upfront proton radiation treatment across all tumor histologies (p=0.05). Patients from the EUR superpopulation were significantly more likely to receive proton over photon radiation relative to other superpopulations (OR=1.6, p=0.04; **Figure 5B**) while those of the AMR superpopulation were significantly more likely to receive photon radiation (OR=0.55, p=0.04). Higher rates of photon radiation among the AMR superpopulation were also observed among patients with EPN (OR=0.17, p=0.02).

We incorporated clinical trial and radiation therapy data into survival models to determine if treatment was associated with survival outcomes in tumor histologies for which genetic ancestry was a significant predictor of survival. There was no significant effect of treatment on survival for patients with LGG, EPN, or germ cell tumors, indicating that ancestry-related differences in survival could not be attributed to differences in treatment (**Figure S6A-C**). In patients with MB of the SHH subtype, radiation type was a significant predictor of EFS (p=0.01; **Figure S6D**), and patients who were treated with photon radiation had higher rates of events relative to those treated with proton radiation (HR=7, p=0.01). However, the survival difference between patients from AMR versus EUR superpopulations was still significant in this model (HR=6.7, p=0.02), indicating that genetic ancestry predicts EFS in this subtype independent of treatment.

Discussion

In this study, we sought to investigate the influence of genetic ancestry on prevalence of CNS tumor subtypes, treatment access, and survival outcomes in pediatric patients. Our findings revealed significant associations between predicted genetic ancestry superpopulations, prevalence of CNS tumor histologies and molecular subtypes, patient OS and EFS and upfront treatment approaches. Notably, we also identified genetic ancestry-specific enrichment of *BRAF* fusion breakpoints among patients with LGG. This work provides a crucial framework for better understanding the contributions of intrinsic germline and tumor genetics and societal components of race and ethnicity in pediatric patients with CNS tumors.

Race and ethnicity are social constructs based on membership in a group sharing cultural and behavioral traits, whereas genetic ancestry is based on variations in genomic structure between groups from similar geographic regions[11,23,24]. The concordance between reported race and ethnicity and predicted genetic ancestry can vary significantly by group[11]. In our work, we observed a significant non-random association between predicted ancestry superpopulation and reported race and ethnicity, consistent with previous findings showing high agreement between reported Black and White race and predicted African and European ancestry, respectively, with larger variation between those of reported American Indian, Alaska Native, or Asian race[11].

We found significant associations between predicted genetic ancestry group and cancer group. Patients with *MYC* subtype ATRT or meningioma tumors were significantly enriched among those of predicted AFR ancestry. While previous work has shown an increased incidence of meningiomas among patients of reported Black race[25], our study is the first to report differences in ATRT incidence by genetic ancestry. Ostrom et al. previously investigated race and reported Hispanic ethnicity in relation to ATRT prevalence, but reported no differences in incidence[26]. Our study also showed that patients with oligodendroglioma were enriched among those of predicted EUR superpopulation, aligning with prior studies reporting higher incidence of oligodendroglioma in those of reported White race[25]. We observed significant enrichment of DIPG or DMG tumors among predicted AMR ancestry patients, although this was shown to be driven by inclusion of

patients from PNOC trials that were disproportionately of AMR ancestry. Additionally, germ cell tumors and schwannomas were significantly enriched in patients of predicted EAS and SAS ancestry, respectively.

We next report associations between predicted ancestry superpopulation and tumor molecular subtypes. EUR superpopulation patients with LGG diagnoses had tumors enriched with *BRAF* V600E mutations or non-*BRAF* altered subtypes associated with higher recurrence rates and worse progression-free survival[19]. EAS superpopulation patients with EPN were enriched with *ZFTA* fusion-positive tumors, which are linked with poorer prognosis[27,28]. AFR superpopulation patients with ATRT were enriched with *MYC* subtype tumors, which are associated with worse survival compared to the *TYR* subtype[29].

In our analysis of LGG breakpoint type frequency, we found that 93.6% of patients harbored one of the four most common breakpoints (i.e., 16:09, 15:09, 16:11, or 18:10)[20]. Patients of predicted EAS and SAS ancestry had tumors significantly enriched for the common 15:09 breakpoint, while AMR patients had tumors significantly enriched for rare *KIAA1549::BRAF* fusion breakpoints. Previous work has shown that rare *KIAA1549::BRAF* fusion breakpoints are associated with worse clinical outcomes in patients with LGG[20]. We recapitulate and extend these findings, showing that rare *KIAA1549::BRAF* breakpoints tumors are enriched in anatomic locations that are inherently more difficult to resect (suprasellar and optic pathway). Interestingly, we also observed enrichment of tumors with the common 15:09 *KIAA1549::BRAF* breakpoint in regions less likely to be resected (spinal and suprasellar regions); however, 15:09 breakpoint tumors were not associated with worse EFS relative to other common breakpoint tumors. Lastly, both 15:09 and rare *KIAA1549::BRAF* breakpoints were significantly more likely to be diagnosed as PMAs, which have been shown to exhibit worse progression-free and overall survival relative to PAs[30].

In this cohort, we found that patients of predicted non-European ancestry exhibited significantly worse OS and EFS compared to patients of predicted European ancestry in certain

tumor histologies and molecular subtypes. Patients of non-EUR ancestry with GCT had worse EFS compared to EUR patients. AMR superpopulation patients with SHH medulloblastomas had significantly worse EFS compared to EUR patients. AFR superpopulation patients with LGG or EPN showed significantly worse OS compared to EUR patients, consistent with prior findings of lower survival among Black children with these tumors[31,32]. Furthermore, these observations remained when restricting analyses to specific molecular subtypes (non-*BRAF* altered LGG and *ZFTA* fusion-positive EPN, respectively).

Lastly, we observed differences in treatment approaches by predicted ancestry superpopulation as a key variable linked to survival. AMR and EAS patients with LGG were more likely to receive only a tumor biopsy rather than a partial or near-total gross resection. This was partly attributable to differences in LGG tumor anatomic location among superpopulations, with tumors of AMR patients enriched in regions of mixed anatomical location and those of EAS patients enriched in optic pathway and suprasellar regions. Future studies with larger historically marginalized superpopulation cohorts should assess the true anatomical distribution differences of LGGs across genetic ancestry groups.

Our study also found varying upfront clinical trial enrollment frequencies by genetic ancestry superpopulation. This is encouraging, as it highlights the importance of including patients from diverse genetic and ancestral backgrounds in clinical research and may represent contemporary changes that better facilitate equitable enrollment in trials and research activities[33,34]. In contrast, EUR superpopulation patients were significantly more likely to receive proton over photon radiation, while the converse was observed among AMR patients. This result aligns with a previous study assessing frequency of proton vs. photon radiation therapy in patients with pediatric tumors, which reported higher incidence of proton radiation therapy among non-Hispanic White patients relative to historically marginalized race groups[6]. While radiation type was not consistently associated with survival differences in our cohort, photon radiotherapy has been associated with higher risk of adverse side effects relative to proton radiotherapy[35]. This work

emphasizes the need for assessing frequency of other events that may arise due to disparities in treatment.

Although our work included patients from a diverse set of genetic and predicted ancestral backgrounds, our data came from patients within the United States. Thus, our cohort was significantly enriched for the European superpopulation, with limited numbers of historically marginalized racial and ethnic groups in our study (i.e., AI/AN and NHPI). Expanding access to international pediatric datasets with potentially unique SDoH considerations will be critical for including diverse representation in future research.

While the inclusion of molecular subtypes in our analyses allowed for a more nuanced exploration of tumor risk across ancestry superpopulation, our sample sizes for several subtypes were limited. Further, the discovery of novel molecular features and further subdivision of current molecular subtypes may lead to reduced power to detect ancestry associated prevalence and outcomes. For example, the *SHH* subgroup of medulloblastoma is now categorized into four subtypes based on molecular features and are associated with different survival outcomes[36]. In addition, pathogenic germline variants are known to play a role in the development of CNS tumors and may be enriched in certain genetic superpopulations[34,37]. Epigenetic processes are also known to contribute to cancer development, and differential DNA methylation patterns among racial groups at birth, notably in cancer pathway genes, has previously been reported[38]. However, we did not explore the contribution of germline findings and DNA methylation patterns to observed superpopulation differences in this study. Our data too may be affected by locoregional population enrichment in catchment areas of our CBTN/PNOC enrolling sites. As more data become available through CBTN, PNOC, and other consortia, we expect to expand our analyses to include additional patients, integrate germline and somatic variation, and further assess genetic subpopulations.

Importantly, our work did not consider the contributions of social health determinants to survival, such as socioeconomic status, insurance status, and time to diagnosis, which are well-

known to impact survival[2,39,40]. Although information on these key SDoH was not available for our retrospective analysis, the CBTN and PNOC aim to include these factors, such as the childhood opportunity index or area deprivation index in future prospective and longitudinal data collection. Historically, there has been a lack of reporting of race and ethnicity in clinical trials and we strive to improve this in the pediatric cancer field, as it is essential if we aim to understand clinically-relevant associations with race and/or ethnicity. Future work should explore these societal factors alongside genetic contributions to cancer incidence and survival, as well as access to clinical trials and treatment regimens. While we used reported race and ethnicity retrospectively collected from electronic medical record abstraction, prospective data collection of demographic information using a patient's self-report may more accurately reflect patients' identities.

Our work revealed several new findings in pediatric CNS tumor histologies and molecular subtypes across predicted genetic superpopulations, highlighting associations with survival and upfront treatment approaches. To further improve equity in care and outcomes for children with CNS malignancies, additional research is needed to delineate the extent to which race and ethnicity differences are driven by societal determinants or tumor biology and molecular subtypes related to genetic ancestry.

Data and Code Availability

Raw data for the CBTN and PNOC can be accessed at dbGaP accession number phs002517 or data access request to CBTN. All data and code used to perform analyses and generate figures for this manuscript can be found at https://github.com/d3b-center/pbta-ancestry.

Acknowledgements and Funding

We have deep gratitude to the patients and families who contributed tissue and data through participation in studies by the CBTN and PNOC. C.K. is the recipient of the Robert A. Winn Diversity in Clinical Trials Career Development Award, established by Bristol Myers Squibb

Foundation. This work was also supported in part by the CBTN, the Chad Tough Foundation, and the Division of Neurosurgery at Children's Hospital of Philadelphia.

Conflict of Interest

Angela J. Waanders is a consultant for Alexion and for DayOne Biopharmaceuticals.

Author Contributions

Conceptualization: CK, JLR, RJC, SMW

Methodology: RJC, JLR, MAB, CK

Software: MAB, RJC, CB

Validation: RJC, JLR, ZG, AC, ASN

Formal Analysis: BZ, CZ, RJC, JLR

Investigation: RJC, JLR, CK

Resources: CK, JLR, PBS, ACR, SMu, MP

Data Curation: RJC, JLR, CK, SMc, ED, NVK, AP

Writing – Original Draft: RJC, CCG, CK, JLR

Writing - Review & Editing: RJC, CCG, CK, JLR, SMc, TBD, AC, ED, AJW, AP, MAB, SMW

Visualization: RJC, CCG, JLR, CK

Supervision: JLR, CK

Project Administration: JLR, CK

Funding Acquisition: PBS, ACR, CK

Bibliography

1. Nieblas-Bedolla E, Christophers B, Williams JR, Power-Hays A, Jimenez N, Rodriguez A. Racial and ethnic disparities among children with primary central nervous system tumors in the US. *J Neurooncol.* 2021; 152(3):451-466.
2. Fineberg R, Zahedi S, Eguchi M, Hart M, Cockburn M, Green AL. Population-Based Analysis of Demographic and Socioeconomic Disparities in Pediatric CNS Cancer Survival in the United States. *Sci Rep.* 2020; 10(1):4588.
3. Moore KJ, Moertel CL, Williams LA. Minority children experience a higher risk of death from many central nervous system tumor types even after accounting for treatment received: A National Cancer Database analysis. *Cancer.* 2022; 128(8):1605-1615.
4. Siegel DA, Li J, Ding H, Singh SD, King JB, Pollack LA. Racial and ethnic differences in survival of pediatric patients with brain and central nervous system cancer in the United States. *Pediatr Blood Cancer.* 2019; 66(2):e27501.
5. Haizel-Cobbina J, Spector LG, Moertel C, Parsons HM. Racial and ethnic disparities in survival of children with brain and central nervous tumors in the United States. *Pediatr Blood Cancer.* 2021; 68(1):e28738.
6. Bitterman DS, Bona K, Laurie F, et al. Race Disparities in Proton Radiotherapy Use for Cancer Treatment in Patients Enrolled in Children's Oncology Group Trials. *JAMA Oncol.* 2020; 6(9):1465-1468.
7. Mersha TB, Abebe T. Correction to: Self-reported race/ethnicity in the age of genomic research: its potential impact on understanding health disparities. *Hum Genomics.* 2021; 15(1):35.
8. Carrot-Zhang J, Chambwe N, Damrauer JS, et al. Comprehensive Analysis of Genetic Ancestry and Its Molecular Correlates in Cancer. *Cancer Cell.* 2020; 37(5):639-654 e636.
9. Yuan J, Hu Z, Mahal BA, et al. Integrated Analysis of Genetic Ancestry and Genomic Alterations across Cancers. *Cancer Cell.* 2018; 34(4):549-560 e549.
10. Lee SHR, Antillon-Klussmann F, Pei D, et al. Association of Genetic Ancestry With the Molecular Subtypes and Prognosis of Childhood Acute Lymphoblastic Leukemia. *JAMA Oncol.* 2022; 8(3):354-363.
11. Shraga R, Yarnall S, Elango S, et al. Evaluating genetic ancestry and self-reported ethnicity in the context of carrier screening. *BMC Genet.* 2017; 18(1):99.
12. Shapiro JA, Gaonkar KS, Spielman SJ, et al. OpenPBTA: The Open Pediatric Brain Tumor Atlas. *Cell Genom.* 2023; 3(7):100340.
13. Geng Z, Wafula E, Zhang Y, Corbett RJ, Jin R, Gaonkar KS, Shukla S, et al. D3b-center/openpedcan-manuscript: Release V1.0.0. https://doi.org/10.5281/zenodo.11496817. Accessed June 6, 2024.
14. Pedersen BS BP, Brown J, et al. Somalier: rapid relatedness estimation for cancer and germline studies using efficient genome sketches. *Genome Med.* 2020; 12(1):62.
15. The 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature.* 2015; 526(7571):68-74.
16. Durinck S, Spellman PT, Birney E, Huber W. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat Protoc.* 2009; 4(8):1184-1191.
17. Durinck S, Moreau Y, Kasprzyk A, et al. BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. *Bioinformatics.* 2005; 21(16):3439-3440.
18. Lawrence M, Huber W, Pages H, et al. Software for computing and annotating genomic ranges. *PLoS Comput Biol.* 2013; 9(8):e1003118.

19. Ryall S, Zapotocky M, Fukuoka K, et al. Integrated Molecular and Clinical Analysis of 1,000 Pediatric Low-Grade Gliomas. *Cancer Cell.* 2020; 37(4):569-583 e565.

20. Haas-Kogan DA, Aboian MS, Minturn JE, et al. Everolimus for Children With Recurrent or Progressive Low-Grade Glioma: Results From the Phase II PNOC001 Trial. *J Clin Oncol.* 2024; 42(4):441-451.

21. A Package for Survival Analysis in R. 2020; https://CRAN.R-project.org/package=survival.

22. survminer: Survival Analysis and Visualization. 2021; https://rpkgs.datanovia.com/survminer/index.html.

23. Kittles RA, Weiss KM. Race, ancestry, and genes: implications for defining disease risk. *Annu Rev Genomics Hum Genet.* 2003; 4:33-67.

24. Cooper RS, Kaufman JS, Ward R. Race and genomics. *N Engl J Med.* 2003; 348(12):1166-1170.

25. Ostrom QT, Gittleman H, Fulop J, et al. CBTRUS Statistical Report: Primary Brain and Central Nervous System Tumors Diagnosed in the United States in 2008-2012. *Neuro Oncol.* 2015; 17 Suppl 4(Suppl 4):iv1-iv62.

26. Ostrom QT, Chen Y, P MdB, et al. The descriptive epidemiology of atypical teratoid/rhabdoid tumors in the United States, 2001-2010. *Neuro Oncol.* 2014; 16(10):1392-1399.

27. Pajtler KW, Witt H, Sill M, et al. Molecular Classification of Ependymal Tumors across All CNS Compartments, Histopathological Grades, and Age Groups. *Cancer Cell.* 2015; 27(5):728-743.

28. Zaytseva M, Papusha L, Novichkova G, Druy A. Molecular Stratification of Childhood Ependymomas as a Basis for Personalized Diagnostics and Treatment. *Cancers (Basel).* 2021; 13(19).

29. Fruhwald MC, Hasselblatt M, Nemes K, et al. Age and DNA methylation subgroup as potential independent risk factors for treatment stratification in children with atypical teratoid/rhabdoid tumors. *Neuro Oncol.* 2020; 22(7):1006-1017.

30. Komotar RJ, Burger PC, Carson BS, et al. Pilocytic and pilomyxoid hypothalamic/chiasmatic astrocytomas. *Neurosurgery.* 2004; 54(1):72-79; discussion 79-80.

31. Stenzel AE, Fenstermaker RA, Wiltsie LM, Moysich KB. Disparities among racial/ethnic groups of patients diagnosed with ependymoma: analyses from the Surveillance, Epidemiology and End Results (SEER) registry. *J Neurooncol.* 2019; 144(1):43-51.

32. Jiang W, Rixiati Y, Kuerban Z, Simayi A, Huang C, Jiao B. Racial/Ethnic Disparities and Survival in Pediatrics with Gliomas Based on the Surveillance, Epidemiology, and End Results Database in the United States. *World Neurosurg.* 2020; 141:e524-e529.

33. Oyer RA, Hurley P, Boehmer L, et al. Increasing Racial and Ethnic Diversity in Cancer Clinical Trials: An American Society of Clinical Oncology and Association of Community Cancer Centers Joint Research Statement. *J Clin Oncol.* 2022; 40(19):2163-2171.

34. Muskens IS, Zhang C, de Smith AJ, Biegel JA, Walsh KM, Wiemels JL. Germline genetic landscape of pediatric central nervous system tumors. *Neuro Oncol.* 2019; 21(11):1376-1388.

35. Baumann BC, Mitra N, Harton JG, et al. Comparative Effectiveness of Proton vs Photon Therapy as Part of Concurrent Chemoradiotherapy for Locally Advanced Cancer. *JAMA Oncol.* 2020; 6(2):237-246.

36. Malbari F, Lindsay H. Genetics of Common Pediatric Brain Tumors. *Pediatr Neurol.* 2020; 104:3-12.

37. Zhang J, Walsh MF, Wu G, et al. Germline Mutations in Predisposition Genes in Pediatric Cancer. *N Engl J Med.* 2015; 373(24):2336-2346.

38. Adkins RM, Krushkal J, Tylavsky FA, Thomas F. Racial differences in gene-specific DNA methylation levels are present at birth. *Birth Defects Res A Clin Mol Teratol.* 2011; 91(8):728-736.

39. Aguirre AO, Lim J, Kuo CC, et al. Social Determinants of Health and Associations With Outcomes in Pediatric Patients With Brain Tumors. *Neurosurgery.* 2023.

40. Puthenpura V, Canavan ME, Poynter JN, et al. Racial/ethnic, socioeconomic, and geographic survival disparities in adolescents and young adults with primary central nervous system tumors. *Pediatr Blood Cancer.* 2021; 68(7):e28970.

Table 1. Frequency of clinical trial enrollment and radiation type by superpopulation and tumor histology

Cancer group	Upfront enrollment in clinical trial						Upfront Proton Radiation						Upfront Photon Radiation					
	AFR	AMR	EAS	EUR	SAS	p	AFR	AMR	EAS	EUR	SAS	p	AFR	AMR	EAS	EUR	SAS	p
Atypical Teratoid Rhabdoid Tumor	3 (27%)	4 (40%)	0	3 (12%)	--	0.23	8 (73%)	6 (60%)	0	15 (58%)	--	0.43	1 (9%)	1 (10%)	1 (100%)	1 (4%)	--	0.14
Choroid plexus tumor	--	0	0	1 (4%)	--	1.00	--	0	0	1 (4%)	--	--	--	--	--	--	--	--
Craniopharyngioma	0	0	0	1 (2%)	0	1.00	4 (33%)	3 (30%)	1 (100%)	6 (15%)	2 (50%)	0.34	0	1 (10%)	0	0	0	0.67
DIPG or DMG	3 (33%)	19 (70%)	1 (20%)	20 (34%)	3 (60%)	***0.01***	0	2 (14%)	0	11 (23%)	0	0.37	4 (50%)	6 (43%)	1 (25%)	17 (36%)	1 (50%)	0.71
Ependymoma	0	6 (38%)	0	10 (13%)	0	***0.04***	9 (75%)	8 (50%)	3 (38%)	46 (58%)	0	0.34	1 (8%)	5 (31%)	0	5 (6%)	0	***0.02***
Germ cell tumor	0	0	1 (33%)	1 (8%)	--	0.65	1 (50%)	2 (67%)	0	8 (67%)	--	1.00	0	0	0	1 (0%)	0	1.00
Low-grade glioma	2 (5%)	4 (9%)	0	19 (7%)	0	0.93	2 (5%)	2 (4%)	0	16 (6%)	0	0.67	0	7 (23%)	2 (15%)	15 (12%)	0	0.38
Medulloblastoma	3 (25%)	8 (27%)	2 (15%)	39 (32%)	1 (33%)	0.78	8 (67%)	17 (57%)	7 (54%)	65 (54%)	2 (67%)	0.90	0	0	1 (33%)	1 (5%)	1 (50%)	0.10
Meningioma	--	--	--	--	--	--	2 (20%)	0	0	4 (20%)	0	1.00	0	1 (14%)	0	0	0	0.35
Mesenchymal tumor	0	0	1 (33%)	2 (7%)	2 (67%)	***0.048***	1 (33%)	1 (14%)	0	13 (45%)	1 (33%)	0.08	0	1 (5%)	0	2 (1%)	1 (17%)	0.46
Mixed neuronal-glial tumor	0	0	0	5 (4%)	0	1.00	0	0	1 (25%)	4 (3%)	0	0.62	0	0	0	1 (2%)	0	--
Neurofibroma Plexiform	0	0	--	1 (9%)	0	1.00	--	--	--	--	--	--	--	--	--	--	--	--
Other CNS embryonal tumor	1 (50%)	2 (50%)	--	1 (10%)	1 (50%)	0.20	1 (50%)	2 (50%)	--	6 (60%)	0	0.39	0	1 (25%)	--	0	0	0.46
Other high-grade glioma	4 (67%)	2 (14%)	3 (43%)	9 (15%)	1 (50%)	***0.01***	2 (33%)	4 (31%)	1 (20%)	23 (40%)	1 (50%)	0.75	2 (33%)	3 (23%)	2 (40%)	11 (19%)	1 (50%)	0.71
Other tumor	0	4 (33%)	0	5 (20%)	--	0.45	2 (33%)	3 (25%)	1 (50%)	8 (32%)	--	1.00	2 (33%)	0	0	2 (8%)	--	0.29
Schwannoma	0	0	0	1 (7%)	0	1.00	0	1 (33%)	0	1 (7%)	0	1.00	0	0	0	1 (7%)	0	1.00
Total	16 (10%)	49 (22%)	8 (12%)	118 (12%)	8 (19%)	***0.001***	40 (66%)	52 (53%)	14 (41%)	227 (62%)	6 (46%)	0.05	10 (16%)	26 (26%)	7 (21%)	58 (16%)	4 (31%)	0.11

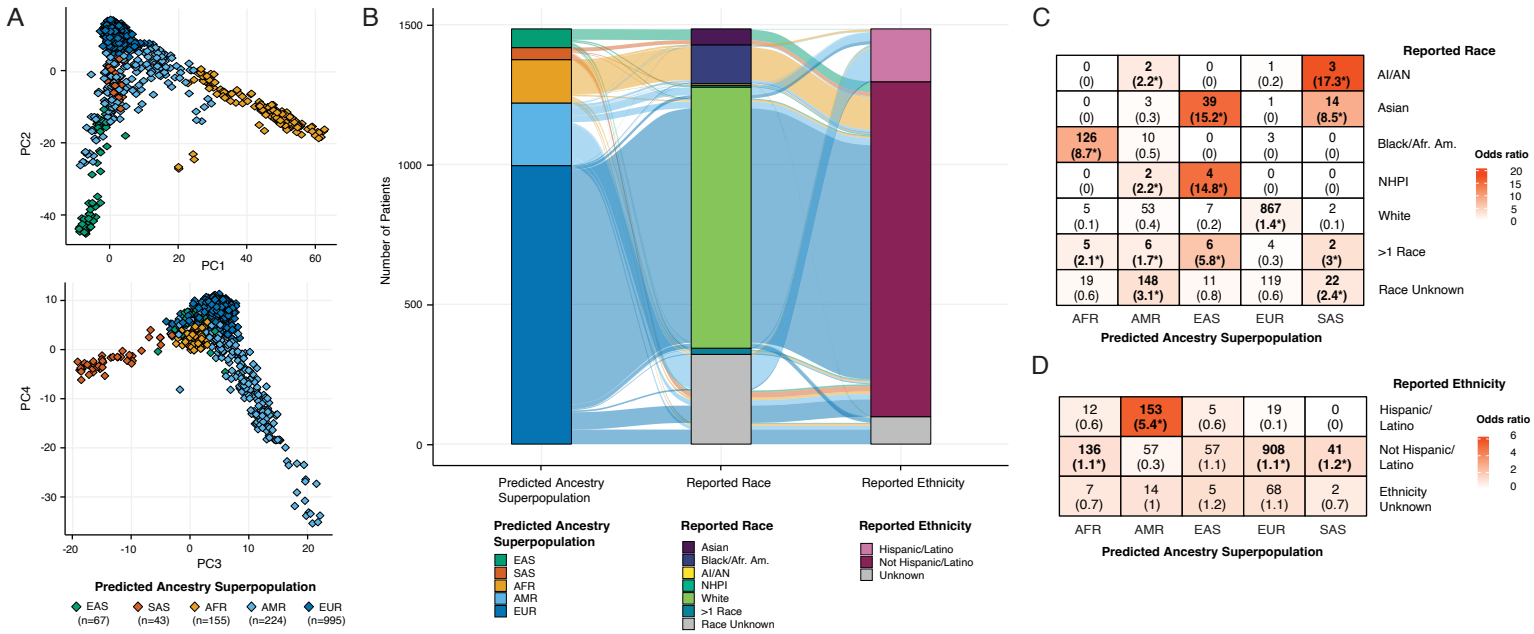

Figure 1. Genetic ancestry prediction in a pediatric central nervous system tumor cohort. A. Somalier genetic ancestry prediction principal component (PC) plot of PCs 1-4. **B.** Alluvial plot displaying relationships between genetic ancestry superpopulation, reported race and ethnicity in pediatric CNS tumor cohort. **C-D**. Enrichment of genetic ancestry superpopulations among each (**C**) reported race and (**D**) ethnicity category. Stars denote Fisher's exact test p<0.05.

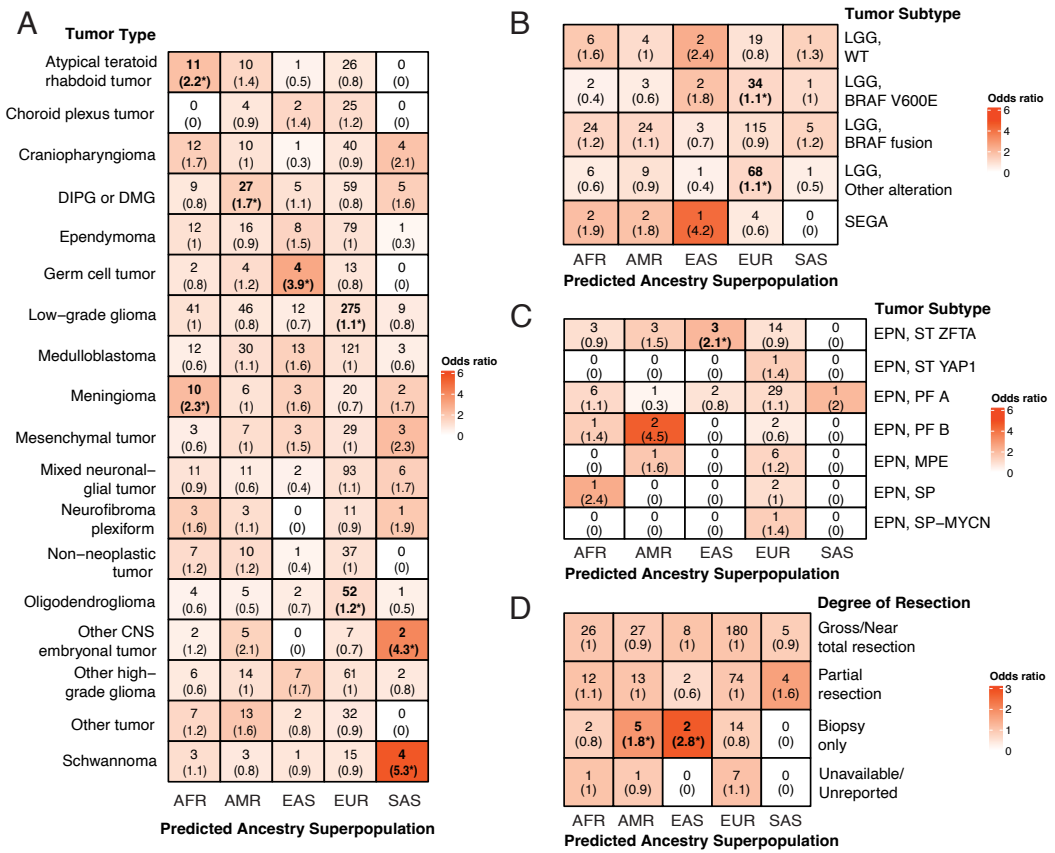

Figure 2. Distribution of CNS tumor histologies, molecular subtypes, and surgical strategies across genetic ancestry superpopulations. A. Count and enrichment of predicted genetic ancestry superpopulations within CNS tumor histologies. *FDR<0.05. **B-C.** Count and enrichment of predicted ancestry superpopulations within LGG (B) and EPN (C) molecular subtype cohorts. **D.** Count and enrichment of predicted genetic ancestry superpopulations for LGG tumor degree of resection. WT=wildtype, EPN=ependymoma, PF=posterior fossa, MPE=myxopapillary ependymoma, ST=supratentorial, SP=spinal. Stars denote Fisher's exact test p<0.05.

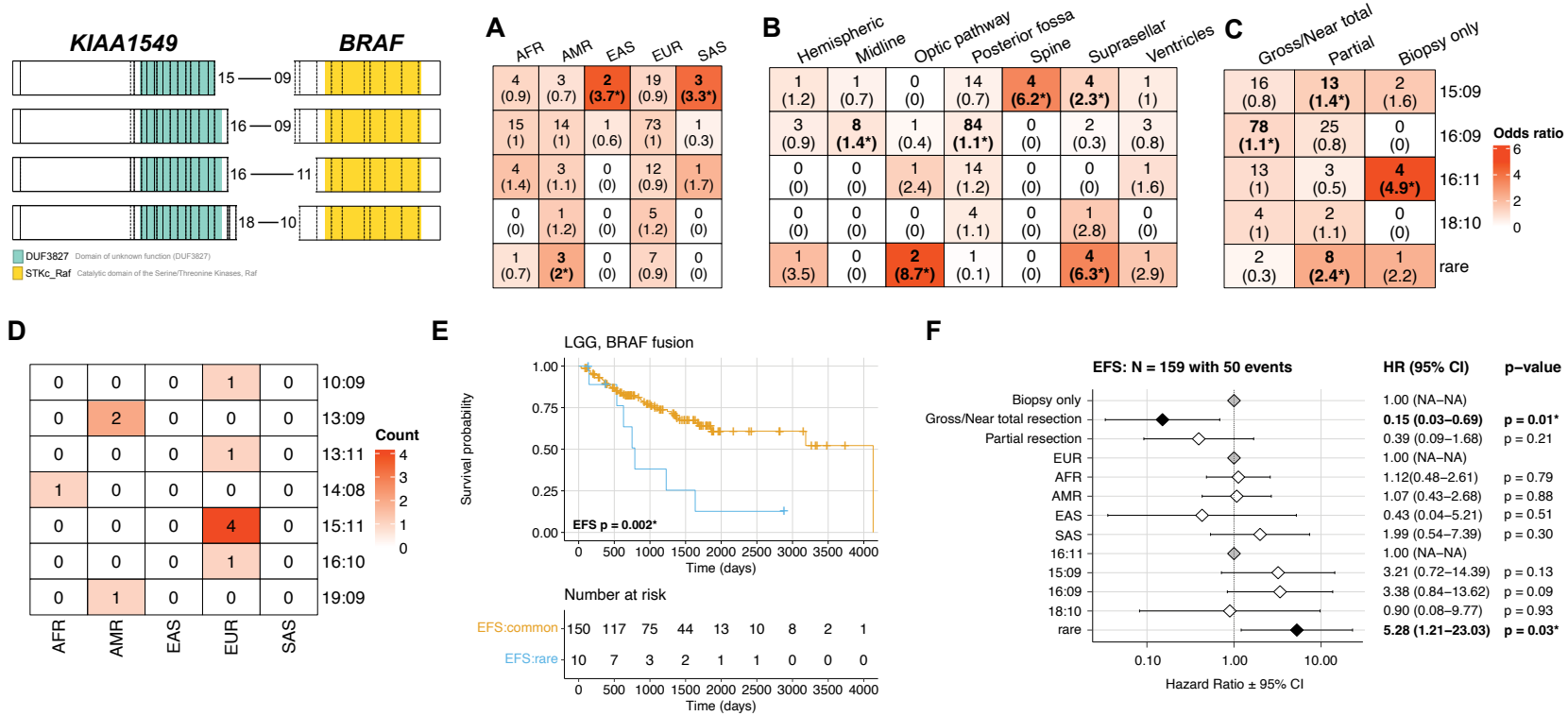

Figure 3. *KIAA1549*::*BRAF* fusion breakpoint distribution by genetic ancestry. A-C. Count and enrichment of *KIAA1549*::*BRAF* fusion breakpoint groups by (**A**) genetic ancestry superpopulation, (**B**) CNS anatomic location, and (**C**) degree of tumor resection. *p<0.05. **D.** Distribution of rare and novel fusion breakpoints by genetic ancestry superpopulation. **E.** Kaplan-Meier event-free survival curves for LGG *BRAF* fusion patients with common and rare breakpoints. **F.** Cox proportional hazards model forest plots of event-free survival in LGG *BRAF* fusion cohort, including covariates for tumor resection level, genetic ancestry superpopulation, and breakpoint type (common vs. rare

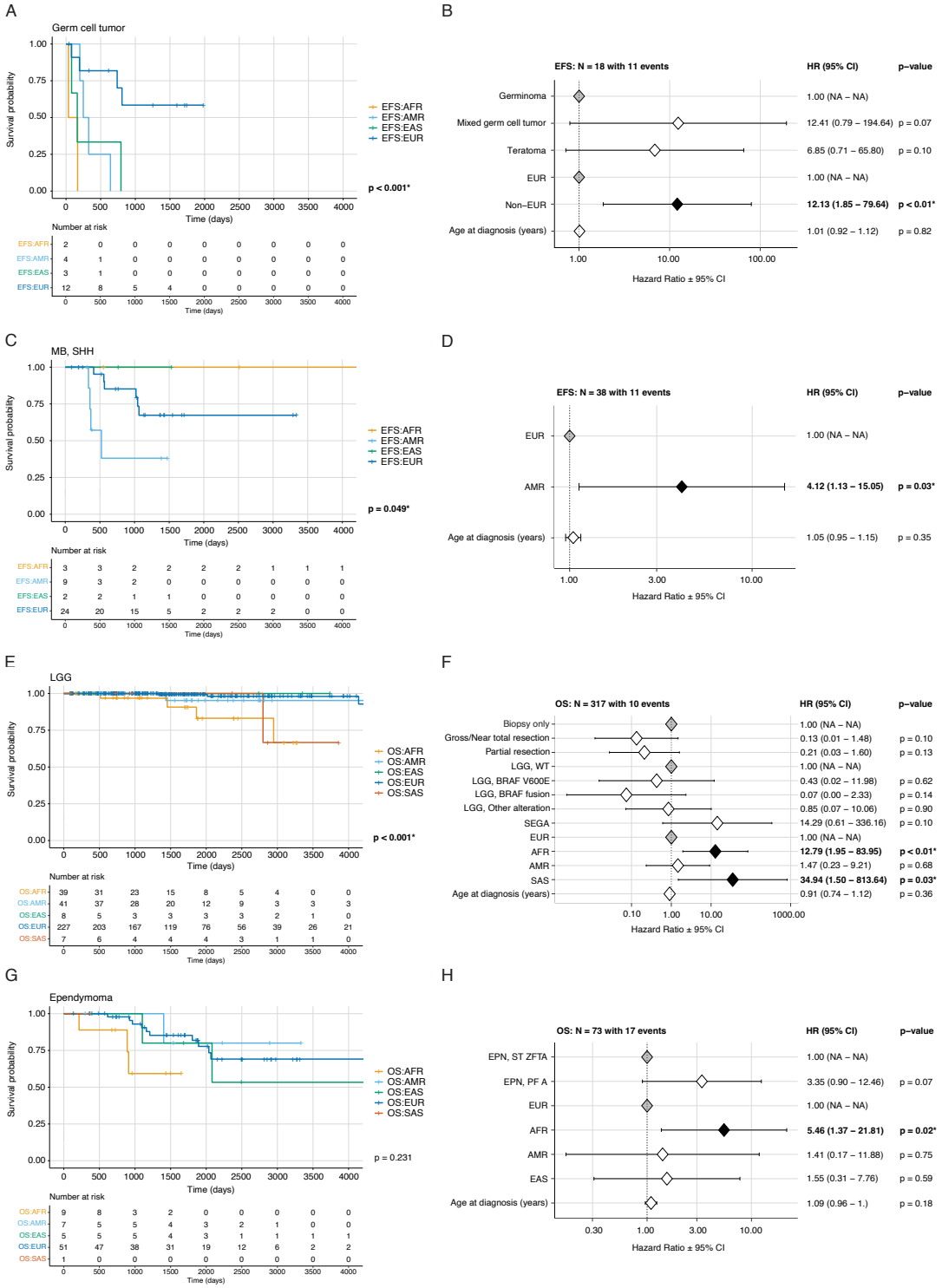

Figure 4. Genetic ancestry-associated overall and event-free survival differences in pediatric CNS tumor patients. A-D. Kaplan-Meier event-free survival curve in patients with **(A)** germ cell tumors by genetic ancestry superpopulation and **(C)** medulloblastoma, *SHH* (MB, *SHH*)

subtype. Cox proportional hazards model forest plots of event-free survival in patients with **(B)** germ cell tumors, including covariates for predicted ancestry, age at diagnosis, and CNS tumor histology (germ cell tumor only) and **(D)** MB, *SHH* subtype. **E-H.** Kaplan-Meier overall survival curve in patients with **(E)** LGG and **(G)** EPN tumors by genetic ancestry superpopulation. Cox proportional hazards model forest plots of OS in patients with **(F)** LGG and **(H)** EPN tumors, including covariates for molecular subtype, genetic ancestry superpopulation, age at diagnosis, and extent of tumor resection (LGG only).

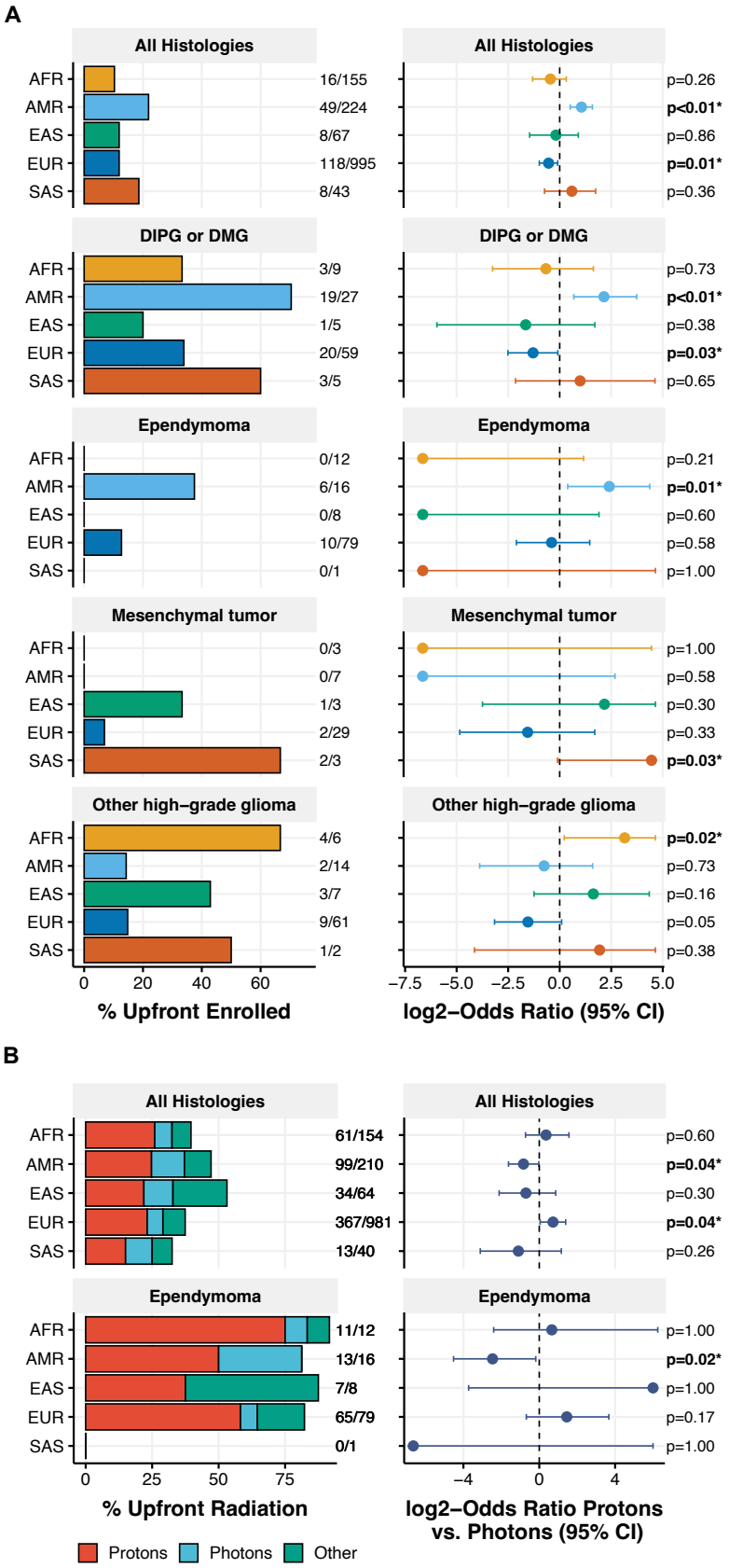
A
All Histologies
AFR
AMR
EAS
EUR
SAS
16/155
49/224
8/67
118/995
8/43
p=0.26
p<0.01*
p=0.86
p=0.01*
p=0.36
DIPG or DMG
3/9
19/27
1/5
20/59
3/5
p=0.73
p<0.01*
p=0.38
p=0.03*
p=0.65
Ependymoma
0/12
6/16
0/8
10/79
0/1
p=0.21
p=0.01*
p=0.60
p=0.58
p=1.00
Mesenchymal tumor
0/3
0/7
1/3
2/29
2/3
p=1.00
p=0.58
p=0.30
p=0.33
p=0.03*
Other high−grade glioma
4/6
2/14
3/7
9/61
1/2
p=0.02*
p=0.73
p=0.16
p=0.05
p=0.38
0
20
40
60
% Upfront Enrolled
−7.5
−5.0
−2.5
0.0
2.5
5.0
log2−Odds Ratio (95% CI)
B
All Histologies
61/154
99/210
34/64
367/981
13/40
p=0.60
p=0.04*
p=0.30
p=0.04*
p=0.26
Ependymoma
11/12
13/16
7/8
65/79
0/1
p=1.00
p=0.02*
p=1.00
p=0.17
p=1.00
0
25
50
75
% Upfront Radiation
−4
0
4
log2−Odds Ratio Protons vs. Photons (95% CI)
Protons
Photons
Other

Figure 5. Pediatric CNS tumor treatment frequency by genetic ancestry superpopulation. **A**. Rates of upfront clinical trial enrollment across all tumor histologies and in specific histologies by genetic ancestry superpopulation, and corresponding odds ratio of enrollment in each superpopulation relative to others. **B.** Rates of radiation therapy treatment by radiation type and genetic ancestry superpopulation across all tumor histologies and in patients with EPN. Odds ratios (ORs) indicate likelihood of receiving proton versus photon radiation in superpopulation relative to others, where $\log_2$-OR > 0 and $\log_2$-OR < 0 indicate significantly increased likelihood of proton and photon radiation, respectively.