Title
European genetic ancestry associated with risk of childhood ependymoma

Journal
Neuro-Oncology, 22(11)

ISSN
1522-8517

Authors
Zhang, Chenan
Ostrom, Quinn T
Hansen, Helen M
et al.

Publication Date
2020-11-26

DOI
10.1093/neuonc/noaa130

Peer reviewed

Neuro-Oncology

22(11), 1637–1646, 2020 | doi:10.1093/neuonc/noaa130 | Advance Access date 2 June 2020

European genetic ancestry associated with risk of childhood ependymoma

Chenan Zhang, Quinn T. Ostrom, Helen M. Hansen, Julio Gonzalez-Maya, Donglei Hu, Elad Ziv, Libby Morimoto, Adam J. de Smith, Ivo S. Muskens, Cassie N. Kline, Zalman Vaksman, Hakon Hakonarson, Sharon J. Diskin, Carol Kruchko, Jill S. Barnholtz-Sloan, Vijay Ramaswamy, Francis Ali-Osman, Melissa L. Bondy, Michael D. Taylor, Catherine Metayer, Joseph L. Wiemels,† and Kyle M. Walsh†

Department of Epidemiology and Biostatistics, University of California San Francisco, San Francisco, California, USA (C.Z., K.M.W.); Department of Medicine, Section of Epidemiology and Population Sciences, Dan L. Duncan Comprehensive Cancer Center, Baylor College of Medicine, Houston, Texas, USA (Q.T.O., M.L.B.); Central Brain Tumor Registry of the United States, Hinsdale, Illinois, USA (Q.T.O., C.K., J.S.B-S.); Department of Neurological Surgery, University of California San Francisco, San Francisco, California, USA (H.M.H., J.G-M.); Division of General Internal Medicine, Department of Medicine, University of California San Francisco, San Francisco, California, USA (D.H., E.Z.); School of Public Health, University of California Berkeley Berkeley, California, USA (L.M., C.M.); Center for Genetic Epidemiology, University of Southern California, Los Angeles, California, USA (A.J.d.S., I.S.M., J.L.W.); Division of Pediatric Hematology/Oncology, Department of Pediatrics, University of California San Francisco, San Francisco, California, USA (C.N.K.); Department of Neurology, University of California San Francisco, San Francisco, California, USA (C.N.K.); Department of Biomedical and Health Informatics, Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, USA (Z.V., S.J.D.); Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, USA (H.H.); Department of Pediatrics, Children's Hospital of Philadelphia and Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA (H.H., S.J.D.); Department of Population and Quantitative Health Sciences and Case Comprehensive Cancer Center, Case Western Reserve University School of Medicine, Cleveland, Ohio, USA (J.S.B-S.); The Arthur and Sonia Labatt Brain Tumor Research Centre, The Hospital for Sick Children, Toronto, Ontario, Canada (V.R., M.D.T.); Department of Neurosurgery and Duke Cancer Institute, Duke University School of Medicine, Durham, North Carolina, USA (F.A-O., K.M.W.)

†Co-senior authors.

Corresponding Author: Kyle M. Walsh, DUMC Box 3050, Durham, NC 27710 (Kyle.Walsh@Duke.edu).

Abstract

Background. Ependymoma is a histologically defined central nervous system tumor most commonly occurring in childhood. Population-level incidence differences by race/ethnicity are observed, with individuals of European ancestry at highest risk. We aimed to determine whether extent of European genetic ancestry is associated with ependymoma risk in US populations.

Methods. In a multi-ethnic study of Californian children (327 cases, 1970 controls), we estimated the proportions of European, African, and Native American ancestry among recently admixed Hispanic and African American subjects and estimated European admixture among non-Hispanic white subjects using genome-wide data. We tested whether genome-wide ancestry differences were associated with ependymoma risk and performed admixture mapping to identify associations with local ancestry. We also evaluated race/ethnicity-stratified ependymoma incidence data from the Central Brain Tumor Registry of the United States (CBTRUS).

Results. CBTRUS data revealed that African American and Native American children have 33% and 36%, respectively, reduced incidence of ependymoma compared with non-Hispanic whites. In genetic analyses, a 20% increase in European ancestry was associated with a 1.31-fold higher odds of ependymoma among self-reported Hispanics and African Americans (95% CI: 1.08–1.59, $P_{meta} = 6.7 \times 10^{-3}$). Additionally, eastern European ancestral substructure was associated with increased ependymoma risk in non-Hispanic whites ($P = 0.030$) and in Hispanics ($P = 0.043$).

Admixture mapping revealed a peak at 20p13 associated with increased local European ancestry, and targeted fine-mapping identified a lead variant at rs6039499 near *RSPO4* (odds ratio = 1.99; 95% CI: 1.45–2.73; $P = 2.2 \times 10^{-5}$) but which was not validated in an independent set of posterior fossa type A patients. **Conclusions.** Interethnic differences in ependymoma risk are recapitulated in the genomic ancestry of ependymoma patients, implicating regions to target in future association studies.

Key Points

1. Childhood ependymoma is less common in African-Americans and Native Americans.
2. Extent of European genomic ancestry increases childhood ependymoma risk.
3. European ancestral substructure is associated with ependymoma risk in non-Hispanic white children.

Importance of the Study

Ependymoma incidence varies across racial/ethnic groups. We analyzed CBTRUS data and observed that African American and Native American children have significantly lower incidence of ependymoma compared with non-Hispanic whites. Although genetic factors have been speculated to contribute in part to these differences, no study to date has investigated this. Our multi-ethnic case-control study of childhood ependymoma found that greater genetically estimated European ancestry was associated with higher ependymoma risk among recently admixed Hispanic and African American children. Furthermore, eastern European ancestral substructure was associated with increased ependymoma risk in non-Hispanic whites and in Hispanics.

Ependymoma is a histologically defined central nervous system tumor most commonly occurring in childhood, accounting for 5.1% of childhood brain tumors.[1] Similar to many childhood brain tumors,[2] little is known regarding factors predisposing individuals to ependymoma risk aside from neurofibromatosis 2.[3–5] Germline mutations in adenomatous polyposis coli, neurofibromatosis type 1 (*NF1*), and tumor protein p53 genes have also been linked to ependymoma,[6–8] although a recent sequencing study including ependymoma patients did not reveal any likely pathogenic germline mutations in known cancer predisposition genes.[9] Population-based epidemiologic studies suggest that incidence rates differ across the world, with the highest rates in countries with individuals of predominantly European ancestry.[10] These racial and ethnic differences are observed in the US, with higher rates in non-Hispanic white individuals compared with other populations.[1] Potential explanations for these incidence differences include varying underlying genetic or environmental risk factors and, potentially, unequal access to medical imaging.

Although racial and ethnic differences in childhood ependymoma incidence are recognized, they have not been rigorously evaluated by tumor location or across histopathological subgroups. Furthermore, whether these incidence differences can be attributed to genetic ancestry or to specific genetic loci has not been explored. To address this, we re-analyzed population-based registry data from the Central Brain Tumor Registry of the United States (CBTRUS) to explore racial and ethnic differences in childhood ependymoma incidence, including variation by histopathological subgroup and tumor location. We next estimated genetic ancestry in a multi-ethnic population of California children with ependymoma and controls to assess associations between ancestral fractions and ependymoma risk. Furthermore, we conducted admixture mapping to examine the association between local European ancestry and ependymoma risk. In addition to clarifying the role of genetic factors contributing to racial/ethnic differences in ependymoma risk, these results can inform future ependymoma genomic research by implicating candidate genes and pathways for further exploration.

Materials and Methods

CBTRUS Data Extraction Methods

Incidence data for ependymal tumors were obtained from CBTRUS, which includes incidence data from 100% of the US population[11,12] and is obtained through a data release agreement with the Centers for Disease Control and Prevention's National Program of Cancer Registries. Average annual incidence rates with 95% confidence intervals were generated for children 0–19 years old at time of diagnosis from 2000–2015, by race/ethnicity, tumor site, and specific ependymal histologies based on International Classification of Disease for Oncology (ICD-O-3) site, histology, and behavior codes using SEER*Stat

software version 8.3.5, as previously described.[13] All rates are standardized to the 2000 US population and reported per 100000 population. Incidence rate ratios were generated to compare incidence rates between racial/ethnic groups.[14,15]This study was approved by the University Hospitals Cleveland Medical Center Institutional Review Board.

Study Participants

Since 1982, newborn blood samples from neonates born within California have been collected by the California Department of Public Health, Genetic Diseases Screening Branch, for the purpose of disease screening, with remaining samples archived at –20°C and made available for approved research. We linked statewide birth records for the years 1982–2009 to data from the California Cancer Registry (CCR) for diagnosis years 1988–2011. Included in this analysis are 327 ependymoma cases and 1970 controls for whom we were able to retrieve a dried newborn bloodspot, representing a 95.1% retrieval rate. Cases were defined as those diagnosed with ependymoma before age 20, per CCR record of ICD-O-3 codes 9391–9394. Histology was coded directly for cases diagnosed 2001 and forward, while cases coded prior to 2001 were forward-converted to ICD-O-3 codes by the CCR. All ependymoma cases in the subset that underwent genomic analyses of bloodspots had a microscopic confirmation of ependymoma (ie, tissue diagnosis). CCR data and CBTRUS data were aligned in terms of ICD-O codes to facilitate comparisons.

Controls were defined as children born in California during the same time period as the cases (1982–2009) and not reported to CCR as having childhood cancer. We ensured that the distribution of cases and controls was similar across years of birth via both graphical plots and Kolmogorov–Smirnov tests. All subjects were assigned to analytic subgroups based on self-reported race/ethnicity, abstracted from birth records data. The study was approved by the institutional review boards at the University of California Berkeley, the University of California San Francisco, the California Department of Public Health, the Hospital for Sick Children (Sick Kids), and the Children's Hospital of Pennsylvania (CHOP). Details on the linkage and use of neonatal bloodspots for studying pediatric cancers have been reported previously.[16]

DNA Extraction and Genotyping

DNA was extracted from a one-third portion of a 12-mm dried bloodspot using the QIAamp DNA Investigator Kit (Qiagen), followed by addition of 280 µL of Buffer ATL and 20 µL of Proteinase K to each sample. Samples were vortexed and then incubated in a dry-bath shaker at 900 rpm and 56°C for one hour. Samples were then briefly centrifuged, after which the lysate solution was transferred to a new 2 mL microcentrifuge tube, and the solid remnants discarded. One microliter of 1 ng/µL carrier RNA was added to the lysate, briefly vortexed, and placed in the Qiagen Qiacube automated work station for DNA isolation, yielding a purified DNA sample in ATE buffer.

DNA was genotyped on the Affymetrix Axiom World Array (LAT), followed by quality control procedures as previously described.[16] In brief, we performed iterative call-rate filtering for single nucleotide polymorphisms (SNPs) and samples by removing SNPs with call rates <92%, then samples with call rates <95%, then SNPs with call rates <97%, then samples with call rates <97%. Any SNP displaying significant departure from Hardy–Weinberg equilibrium $P < 1.0 \times 10^{-5}$ among non-Hispanic white controls was excluded. Samples with mismatched reported versus genotyped sex were also excluded. Identity-by-descent (IBD) analyses were performed in PLINK on cases and controls,[17,18] with exclusion of one member of any sample pair that had an IBD proportion >0.18. Using genome-wide SNP array data from HapMap Phase 3 samples, we removed any sample showing evidence of mismatched ancestry, ie, >3 SDs from mean MXL, CEU, ASW, or EAS values on the first 3 principal components (PCs). Haplotype phasing was performed with SHAPEIT v2.79029 and whole-genome imputation was carried out using the Minimac3 software[19] with the 2016 release of the Haplotype Reference Consortium.[20] Excluded were SNPs with imputation quality (info) scores <0.60, posterior probabilities <0.90, and minor allele frequencies <0.01 in controls.

Estimation of Ancestry Proportions

We used the unsupervised algorithm in ADMIXTURE, assuming 3 ancestral populations (K = 3), to estimate the proportions of European, African, and Native American ancestry present among self-reported Hispanic and African American individuals. We ran the cross-validation procedure in ADMIXTURE with 10-fold cross-validation for K values 1–6, observing the minimum cross-validation error to occur at K = 3. ADMIXTURE estimates global genetic ancestry using large autosomal SNP genotype datasets from unrelated individuals and a likelihood-based modeling approach.[21] Using a combined dataset of reference individuals from the Human Genome Diversity Project (HGDP)[22] and ependymoma cases and controls (either Hispanic or African American), we first performed linkage disequilibrium (LD) pruning in PLINK with a 50 kb sliding window, 10 kb window shift, and $R^2 < 0.1$, resulting in a starting set of 149996 markers that was thinned to 58441 markers for ancestry estimation. We repeated estimation using fastSTRUCTURE,[23] a Bayesian model-based clustering method, for multiple choices of K ranging 1–6. We observed that the marginal likelihood was maximized at K = 3. We further confirmed that individuals of known parental populations from the reference panel, including Europe ($n = 153$), Africa ($n = 107$), and the Americas ($n = 107$), were estimated as having mean >95% of the correct corresponding ancestry proportions determined from the genetic ancestry algorithms. Among non-Hispanic white individuals, unsupervised estimates of genetic ancestry fractions were also assessed for association with ependymoma risk. Finally, we compared estimated ancestry proportions with eigenvalues from PC analyses to explore correlation between the 2 measures of population structure.

Estimation of Local Ancestry

We used phased data of HGDP subjects with European, African, and Native American ancestry as the reference

dataset to infer local ancestry with the Local Ancestry in admixed Populations using Linkage Disequilibrium (LAMP-LD) method (window size = 200 SNPs, number of states = 15).[24] Local European ancestry was coded as the number of European-ancestry alleles at each SNP (ie, 0, 1, or 2).

Statistical Analysis

We tested for associations between European ancestry and ependymoma risk among self-identified Hispanic subjects and African American subjects using logistic regression models adjusting for sex. We also performed case-only analyses of clinical features, including the association between European ancestry and age at diagnosis, tumor histology, grade, and tumor location.

For local admixture analysis, we tested for associations between local European ancestry at each locus and ependymoma risk among self-identified Hispanic subjects, using logistic regression models adjusting for sex and global genetic ancestry. A strict Bonferroni correction for multiple testing was applied for the 58441 SNPs assessed ($P_{strict} < 8.6 \times 10^{-7}$). We also calculated a suggestive significance threshold using the autocorrelation method proposed by Shriner et al, which estimates the effective number of tests for admixture mapping based on the number of ancestry switches summed across autosomes, averaged across individuals.[25] The effective number of tests for admixture mapping was 67.1 for Hispanics, resulting in a suggestive significance threshold of $P_{suggestive} < 7.5 \times 10^{-3}$.

We then performed association analyses for SNPs within 500 kb of admixture mapping peaks using logistic regression models for Hispanics, non-Hispanic whites, and African Americans separately, adjusting for sex and global genetic ancestry. Meta-analysis was performed to obtain association estimates across the 3 racial/ethnic groups. We calculated the effective number of tests performed in the region, after adjusting for LD between SNPs, and used this number to apply a Bonferroni correction for allelic associations.[26]

Toronto Case-Control Analyses

A total of 83 ependymoma patients and 332 controls of non-Hispanic white ethnicity were genotyped on the Illumina OmniExpress array at CHOP. All ependymoma patients were recruited onto study at Sick Kids and had posterior fossa type A (PF-EPN-A) tumors.[27] DNA was extracted from blood where available, but a subset of DNA samples were extracted from tumor specimen. Because PF-EPN-A are genetically bland and rarely harbor copy-number alterations, negative impacts on genotyping quality appeared minimal.[27] Data from genome-wide area studies (GWASes) underwent quality control procedures as previously described.[28] From a control group of 3254 children of European ancestry who were recruited and genotyped at CHOP, genome-wide identity-by-state (IBS) estimates for all pairwise comparisons of case and control subjects were calculated and 4 controls were selected per case based on nearest-neighbor matching using genome-wide IBS estimates, yielding 332 controls matched to our 83 cases. The 4 nearest controls were matched to each case to correct for population structure, as previously described.[28] Genotypes were phased using SHAPEIT2,[29] and whole-genome imputation was performed using IMPUTE2,[30,31] with 1000 Genomes Phase 3 release as the imputation reference panel. Because the external controls were tightly matched to cases on genomic ancestry and IBS, ancestry-based case-control comparisons were not appropriate. However, SNP-based case-control comparisons could be made to replicate allelic associations, as implemented in SNPTEST v2.4.1 with adjustment for the first 5 PCs.

Results

CBTRUS Data

Average annual incidence rates (IRs) for ependymal tumors in children 0–19 years old (2000–2015) in the US are presented in Table 1. Rates varied between populations, with the highest rates observed in non-Hispanic whites (IR = 0.30 per 100000 population, 95% CI: 0.28–0.31) and Hispanics (IR = 0.28 per 100000 population, 95% CI: 0.26–0.31). Compared with non-Hispanic whites, American Indian/Alaska Natives were at significantly reduced risk (relative risk [RR] = 0.64, 95% CI: 0.46–0.87, $P = 2.8 \times 10^{-3}$), as were African Americans (RR = 0.67, 95% CI: 0.60–0.74, $P = 4.1 \times 10^{-14}$). Asian/Pacific Islanders also had a lower rate (RR = 0.86, 95% CI: 0.73–1.00, $P = 0.051$). Although no significant differences were observed in overall ependymoma rates among Hispanic individuals compared with non-Hispanic whites (RR = 0.96, 95% CI: 0.88–1.05, $P = 0.367$), lower incidence rates were observed among Hispanics and African Americans for specific histological groupings

Table 1 Annual incidence rates per 100000 persons of ependymal tumors in children ages 0–19, by race/ethnicity (CBTRUS, 2000–2015)

Race/Ethnicity	Count	Incidence Rate (95% CI)	Rate Ratio (95% CI) (compared with non-Hispanic whites)	Ratio *P*-value (compared with non-Hispanic whites)
Non-Hispanic white	2214	0.30 (0.28–0.31)	1	ref
Hispanic white	726	0.28 (0.26–0.31)	0.96 (0.88–1.05)	0.3673
African American	426	0.20 (0.18–0.22)	0.67 (0.60–0.74)	4.1×10^{-14}
American Indian/Alaska Native	42	0.19 (0.14–0.26)	0.64 (0.46–0.87)	2.8×10^{-3}
Asian or Pacific Islander	176	0.25 (0.22–0.29)	0.86 (0.73–1.00)	0.051

Table 2 Demographics and clinical features of 327 multi-ethnic pediatric ependymoma cases (ages 0–19) from the California Department of Public Health

Variable	Median (interquartile range) or *N* (%)
Age at diagnosis	4 (8)
Sex, male	187 (57.2%)
Race/ethnicity (self-reported)	
Non-Hispanic white	153 (46.8%)
Hispanic	142 (43.4%)
African American	32 (9.8%)
Histology	
Anaplastic ependymoma	106 (32.4%)
Ependymoma (NOS)	189 (57.8%)
Myxopapillary	27 (8.3%)
Subependymoma	3 (0.9 %)
Papillary	2 (0.6%)
Location	
Spinal	49 (15.0%)
Intracranial	278 (85%)
World Health Organization Grade	
I	11 (3.4%)
II	31 (9.5%)
III	17 (5.2%)
IV	94 (28.7%)
Not otherwise specified (NOS)	174 (53.2%)

and tumor sites, including myxopapillary ependymomas (Supplementary Table 1) and tumors located in the spine and cauda equina (Supplementary Table 2).

Genome-Wide Ancestry in Hispanics and African Americans

From the California dataset, a total of 142 Hispanic ependymoma patients and 1147 controls, 32 African American patients and 116 controls, and 153 non-Hispanic white patients and 696 controls had data available for genomic analyses (Table 2). Mean composition of Native American, African, and European genetic ancestry fractions, estimated using genomic data, are presented in Table 3 and Supplementary Table 3, and the distribution of ancestry fractions are shown for both case-control samples and the HGDP reference individuals in Fig. 1 Mean values of genetic proportions were highly correlated between ADMIXTURE and fastSTRUCTURE ($R^2 > 0.99$), and ADMIXTURE estimates are reported hereafter. Meta-analysis of ancestry associations across Hispanic and African American subjects identified an association between European ancestry and ependymoma risk, with each 20% increase in European ancestry associated with a 1.31-fold increase in odds of ependymoma (95% CI: 1.08–1.59, $P_{meta} = 6.7 \times 10^{-3}$).

Among Hispanic individuals, European ancestry was associated with 1.34-fold increased odds of ependymoma per 20% increase in European ancestry (95% CI: 1.09–1.67; $P = 6.2 \times 10^{-3}$). Conversely, a greater fraction of Native American ancestry was associated with reduced risk (OR = 0.79 per 20% increase in Native American ancestry, 95% CI: 0.64–0.97; $P = 0.023$). Association analyses between PCs and ependymoma risk among Hispanic individuals showed PC2 and PC3 to be significantly associated (Supplementary Figure 1, inset), suggesting that greater European ancestry (ie, higher PC2 or lower PC3) confers increased ependymoma risk (Supplementary Figure 1). While PC2 separated European from African and Native American ancestry and was associated with greater ependymoma risk ($P = 4.6 \times 10^{-3}$), PC3 values were lowest among individuals from the Russia-Caucasus region and were associated with reduced ependymoma risk ($P = 0.043$).

Among African Americans, a nonsignificant association between ancestry fraction and ependymoma risk was observed (OR = 1.14 per 20% increase in European ancestry, 95% CI: 0.68–1.83; $P = 0.61$). Correspondingly, a nonsignificant protective association was observed for African ancestry (OR = 0.80, 95% CI: 0.53–1.24; $P = 0.30$).

Genome-Wide Ancestry in Non-Hispanic Whites

We next analyzed European substructure in non-Hispanic white ependymoma patients and controls. Association analyses between PCs and ependymoma risk among non-Hispanic whites showed PC1 and PC5 to be significantly associated (Supplementary Figure 2, inset). PC1 roughly corresponds to the east-west axis of Europe, with the PC1 association signal driven by Russia-Caucasus ancestry with lower PC1 (more similar to Russia-Caucasus individuals) conferring greater ependymoma risk (Supplementary Figure 2). PC2 roughly corresponds to the north-south axis of Europe and was not significantly associated with ependymoma risk. A significant PC5 association appears to correspond to French-Basque ancestry. Similar analyses in ADMIXTURE, assuming 2 ancestral populations, revealed significantly increased risk of ependymoma for European subpopulation 1 compared with 2 ($P = 0.029$), where subpopulation 1 corresponded to Adyghe individuals from the Russia-Caucasus region (Supplementary Figure 3).

Adjustments for insurance status at birth (government program vs private insurance/self-pay) did not meaningfully alter effect ancestry associations. In case-only analyses, we did not observe any association between European ancestry fraction and ependymoma tumor site, histology, grade, or age at diagnosis in Hispanics or non-Hispanic whites.

Admixture Mapping in Hispanics

Admixture mapping was carried out in the Hispanic case-control set, and genome-wide associations are summarized in Fig. 2A The top admixture mapping signal was in the 20p13 region at rs6040222 (OR = 1.78; 95% CI: 1.30–2.46 per European allele; $P = 4.1 \times 10^{-4}$). This association was significant after Bonferroni correction for 67 effective tests based on the autocorrelation method ($P < 7.5 \times 10^{-3}$), but did not reach the strict significance threshold ($P < 8.6 \times 10^{-7}$).[26] Additional admixture peaks

Table 3 Mean ancestry fraction in ependymoma cases, ages 0–19, and age-matched controls

Self-Reported Race/Ethnicity	Cases/Controls	Genetically Estimated Ancestry Fractions	Cases Mean (SD)	Control Mean (SD)	*P*-value[a]
Hispanic	142/1147	% Native American	41.7 (15.2)	45.0 (16.7)	0.025
		% African	4.6 (2.5)	5.3 (5.1)	0.11
		% European	53.6 (15.6)	49.6 (16.7)	6.2×10^{-3}
African American	32/116	% Native American	6.7 (9.5)	4.7 (7.7)	0.22
		% African	67.1 (17.6)	70.8 (17.7)	0.30
		% European	26.2 (14.5)	24.5 (15.8)	0.59

[a]Two-tailed *t*-test.

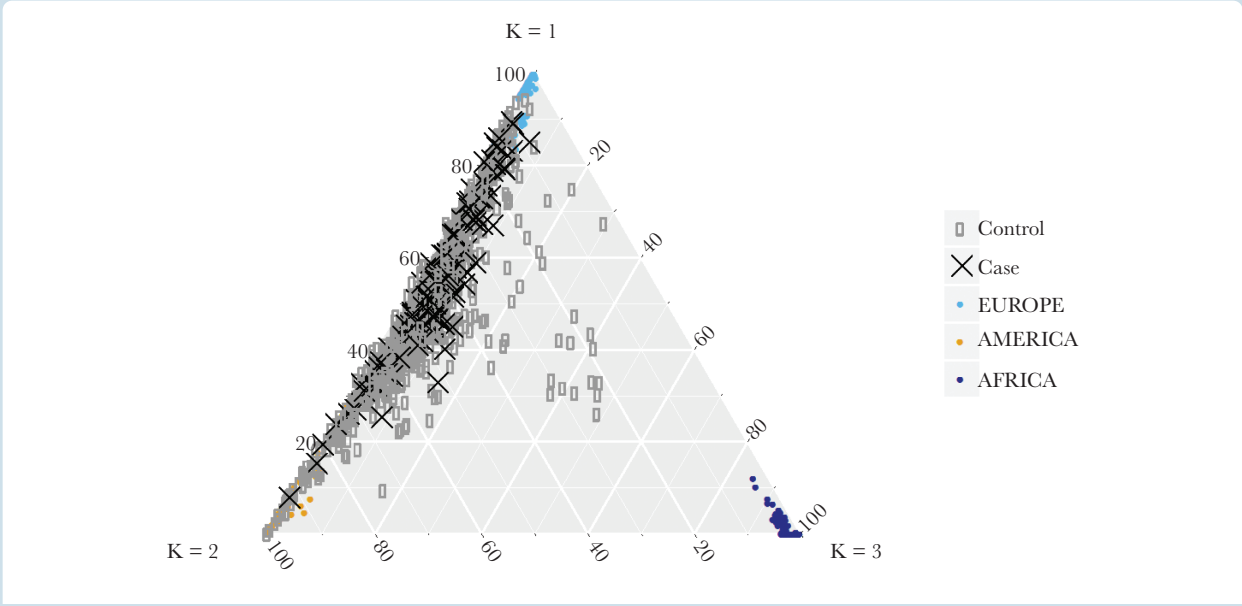

Fig. 1 Triangle plot of ancestry fractions in self-identified Hispanic ependymoma patients (black x's) and controls (gray circles), plotted with reference subjects from the Human Genome Diversity Project. Axes represent fraction of ancestry belonging to each population, with individuals' fractions across 3 populations summing to 1.

with $P < 7.5 \times 10^{-3}$ were located on chromosomes 2p24.3, 2q23.1, 2q36.1, 4q28.2, and 8q24.2. Lead SNPs appear in Supplementary Table 4.

Fine-Mapping Lead Admixture Peak in a Multi-Ethnic Sample

We performed allelic association analyses for 4180 SNPs within 500 kb of the admixture mapping peak at rs6040222 in Hispanics, non-Hispanic whites, and African Americans (Fig. 2B) [32] The most significant association was at rs6039499 (odds ratio [OR] = 1.99; 95% CI: 1.45–2.73; $P = 2.2 \times 10^{-5}$), located 23 kb upstream of the R-spondin-4 (*RSPO4*) gene. The association remained significant after Bonferroni correction for 2089 "effective tests" following adjustment for LD ($P_{corrected}$ = 0.046). Odds ratios for rs6039499 were not meaningfully different when analyses were restricted to intracranial ependymoma (OR = 1.96; 95% CI: 1.40–2.76), spinal ependymoma (OR = 1.98; 95% CI: 1.03–3.81), or posterior fossa ependymoma diagnosed before age 15 (OR = 2.00; 95% CI: 1.39–2.90). The risk (G) allele at rs6039499 is uncommon in European (minor allele frequency [MAF] = 0.06), Native American (MAF = 0.04), and African (MAF = 0.07) subjects from 1000 Genomes, but substantially more common in Asian populations (MAF = 0.28). An adjacent SNP in near complete LD (rs6077564; R^2 = 0.99) is an expression quantitative trait locus for *RSPO4* and *PSMF1* expression in the cerebellum. However, no significant association at rs6039499 was observed among molecularly defined PF-EPN-A patients from Toronto (OR = 0.76, 95% CI: 0.40–1.45, $P = 0.40$).

Discussion

We observed an association between European ancestry and ependymoma risk among admixed Hispanic and African American subjects from California, with each 20%

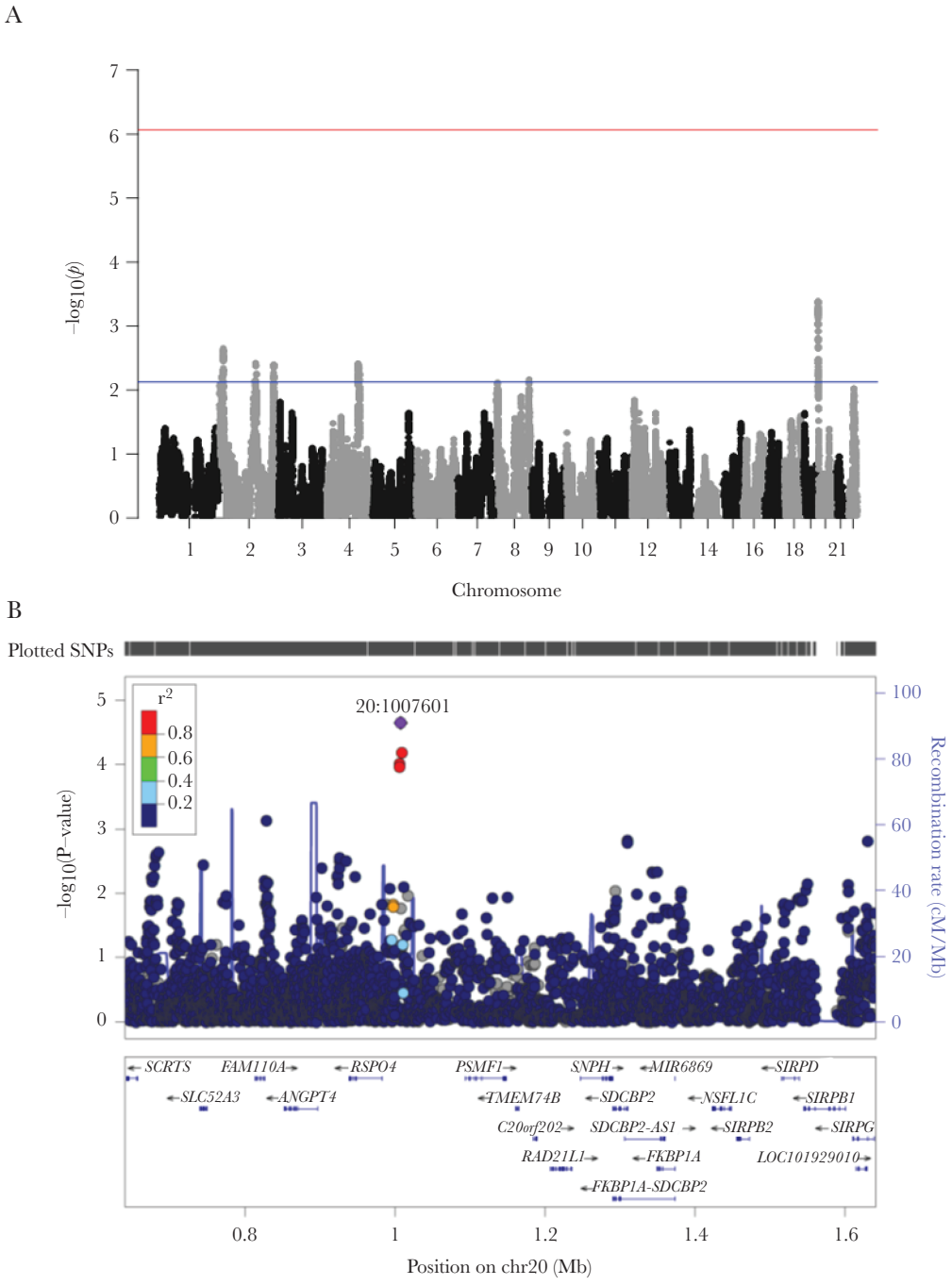

Fig. 2 Association plots of (A) local European ancestry and ependymoma risk among California Hispanics (red and blue lines indicate $P_{strict} < 8.6 \times 10^{-7}$ and $P_{suggestive} < 7.5 \times 10^{-3}$, respectively), and (B) genotype and ependymoma risk in meta-analysis of Hispanics, African Americans, and non-Hispanic whites in a region of chromosome 20 showing the strongest association between local European ancestry and ependymoma risk.

increase in European ancestry associated with a 1.31-fold increased risk of ependymoma. This finding, based on genetic ancestry estimation using genome-wide SNP data with 2 different software packages, was also seen with PC analyses of the top eigenvectors. Importantly, this finding is consistent with epidemiologic observations of greater ependymoma risk among individuals of European ancestry compared with other races/ethnicities, as observed in our analysis of CBTRUS data. Although we did not observe a significant association with European ancestry and ependymoma risk among the African American subset of subjects in this analysis, we may have been underpowered due to small sample size and lower mean European ancestry fraction among African Americans compared with Hispanics. Our mean genetic ancestry estimates are comparable to previous estimates of African (73%), Native

American (1%), and European (24%) proportions among African Americans,[33] as well as that of African (6%), Native American (40%), and European (54%) proportions among Hispanics from California.[34]

Our population-based estimates of higher childhood ependymoma incidence rates in non-Hispanic whites versus other US racial/ethnic groups are consistent with reports for other glial tumors, including adult ependymoma,[35] glioblastoma, lower-grade astrocytoma, and oligodendroglioma.[36] CBTRUS data reveal that the risk of ependymoma in non-Hispanic whites was 50% greater than that in African Americans, which is substantial at the population level. However, given the rarity of this disease, this represents an increase in annual risk from 2 in 1 000 000 to 3 in 1 000 000 at the individual level. Similar patterns have also been noted with pediatric brain tumor diagnoses in patients with NF1.[37] In addition to inherited genetic risk by ancestry, possible causes for the observed incidence differences include ascertainment bias, in which unequal access to care, differences in diagnostic procedures, histologic confirmation, and/or differences in reporting have impacts on incidence rates. However, our findings are unlikely to be driven by such factors thanks to our population-based sampling strategy, >95% bloodspot retrieval rate, and California's opt-out (as opposed to opt-in) research consent. Further, we saw no evidence of differences in tumor grade or age at diagnosis with respect to European ancestry, and adjustment for insurance status at birth did not attenuate the observed associations between ependymoma risk and ancestry.

The possibility that European-ancestry populations may carry a higher frequency of ependymoma risk alleles is intriguing. Interestingly, the Adyghe (Russia-Caucasus) and Basques are two of the more genetically isolated European populations,[38] and greater genetic ancestry fractions corresponding to these two European subpopulations were associated with increased risk of ependymoma. A higher incidence of brain cancer was previously noted in the Navarre and Basque Country regions in an occupational/environmental epidemiology study.[39] Our results suggest that a genetic explanation also merits consideration. Also of note, genetic similarity has previously been described between Adyghe and Ashkenazi Jewish individuals, with the Adyghe population described as the closest European population to Jewish populations.[40,41] It is therefore possible that the signal we describe is partly driven by Ashkenazi Jewish ancestry, but investigating this will necessitate a more targeted approach with reference populations that include a large Ashkenazi sample.

Admixture mapping identified an association peak at 20p13 showing increased European ancestry among ependymoma cases compared with controls, consistent with the global genetic ancestry association results. Follow-up genotypic association analyses in this region identified an association at rs6039499, an intergenic SNP that alters a Hoxa3 transcription factor binding motif. This polymorphism is located 25 kb upstream of *RSPO4*, a regulator of Wnt/beta-catenin signaling, suggesting potential regulatory roles during critical developmental windows. Although SNP associations did not differ by site or histology, we were unable to replicate this association in a molecularly stratified sample of PF-EPN-A from Toronto.

Admixture mapping can serve as a powerful method for identifying candidate loci of interest, particularly in a 2-stage approach.[42] Only one other study has examined the association between genetic ancestry and brain tumor risk. Employing a similar strategy of estimating global and local ancestry using genotyping data, this time in adult glioma patients, Ostrom et al observe a suggestive association with global European ancestry and confirm local ancestry associations at loci previously identified in GWASes of adult glioma.[43] Importantly, admixture mapping is a gene-mapping technique that aims to identify risk alleles that confer disease risk among all individuals harboring the risk allele, not only among those of the particular ancestry groups. Any causal variants underlying the admixture signal has important implications for ependymoma risk across all racial/ethnic groups. Our observation of consistent ancestry associations across strata of ependymoma site and histology, coupled with similar observations for adult glioma, suggest that genomic ancestry may confer risk through an etiologic mechanism that is shared across diverse brain tumor histologies (eg, maintenance methyltransferase activity, glial development).

A primary limitation of our study is a reliance on cancer registry data and lack of molecularly stratified ependymoma subgroups in both the CBTRUS analyses and the genetic analyses of bloodspot specimens linked to California Cancer Registry data. The CBTRUS data used for this analysis cover 100% of the US population, but there is no central pathology review and histology reflects what is assigned by the diagnosing pathologist. In CBTRUS data, race is abstracted from the patient's medical record by cancer registrars, while Hispanic ethnicity is both collected from the chart and assigned via algorithm based on surname. As a result, there may be some error in assignment of histology, race, and/or ethnicity in the CBTRUS results. While molecular subgroups of ependymoma are tightly intertwined with tumor location and we did not observe differences in the effect of ancestry on ependymoma risk when stratifying by tumor site, future analyses incorporating additional molecular subgroup information will help to elucidate the biology underlying the associations observed here.

The lack of environmental covariates also limits our ability to adjust for potential confounders, assess mediating effects, or examine potential gene–environment interactions. However, ependymoma has few known extrinsic risk factors, so our results are unlikely to be strongly confounded. Furthermore, it is unlikely that the same ascertainment biases or confounding factors are simultaneously driving (i) the association between genome-wide European ancestry and ependymoma risk in US Hispanics, (ii) the association between Russia-Caucasus ancestral substructure and ependymoma risk in US Hispanics, and (iii) the association between Russia-Caucasus ancestral substructure and ependymoma risk in US non-Hispanic whites. Additionally, European ancestry proportions were not associated with age at diagnosis in case-only analyses, suggesting that delays in diagnosis are unlikely to confound our results. Therefore, we believe our findings reasonably suggest that population-level differences in childhood ependymoma incidence are, at least partially, driven by a genetic component.

In summary, we demonstrate in an admixed sample of California ependymoma patients and controls that genetic

ancestry contributes to population-level differences in childhood ependymoma incidence across different racial/ethnic groups in the US. Furthermore, our admixture mapping approach has identified candidate genetic regions that future studies can now target, a priori, thereby minimizing the impact of multiple-testing corrections on study power in the setting of uncommon childhood cancers.

Supplementary Material

Supplementary data are available at *Neuro-Oncology* online.

Keywords

disparities | ependymoma | ethnicity | genetic ancestry | pediatric cancer | race

Funding

This work was supported by the Pediatric Brain Tumor Foundation (KMW), The Sontag Foundation (KMW), The National Institutes of Health T32 CA151022-06 (CZ), R01 CA194189 (KMW, JLW), P50 CA190991 (KMW, FAO), R01 CA155461 (JLW), 'A' Awards from The Alex's Lemonade Stand Foundation (AJdS, KMW), and the Cancer Prevention and Research Institute of Texas RP160097T (QTO).

Acknowledgments

Funding for CBTRUS was provided by the Centers for Disease Control and Prevention (CDC) under contract no. 2016-M-9030, the American Brain Tumor Association, The Sontag Foundation, Novocure, Abbvie, and the National Cancer Institute (NCI) under contract No. HHSN261201800176P, the Musella Foundation, the Uncle Kory Foundation as well as private and in kind donations. Contents are solely the responsibility of the authors and do not necessarily reflect the official views of the CDC or the NCI.

The biospecimens and/or data used in this study were obtained from the California Biobank Program (SIS request number 597). The California Department of Public Health is not responsible for the results or conclusions drawn by the authors of this publication. The collection of cancer incidence data used in this study was supported by the California Department of Public Health pursuant to California Health and Safety Code Section 103885; CDC's National Program of Cancer Registries, under cooperative agreement 5NU58DP006344; the National Cancer Institute's Surveillance, Epidemiology, and End Results Program under contract HHSN261201800032I awarded to the University of California San Francisco, contract HHSN261201800015I awarded to the University of Southern California, and contract HHSN261201800009I awarded to the Public Health Institute, Cancer Registry of Greater California. The ideas and opinions expressed herein are those of the author(s) and do not necessarily reflect the opinions of the State of California, Department of Public Health, the NCI, and the CDC or their contractors and subcontractors.

Conflict of interest statement. The authors report no conflicts of interest.

Authorship statement. All authors have read and approved the final version. Additional contributions are as follows: experimental design (CZ, DH, EZ, SJD, VR, MDT, CM, JLW, KMW), implementation (HHM, JG-M, LM, HH, CK, JSB-S, VR, FA-O, MLB, MDT, CM, JLW, KMW), analysis and interpretation of the data (CZ, QTO, HHM, AJdS, CNK, ZV, SJD, JSB-S, MLB, CM, JLW, KMW), writing (CZ, QTO, AJdS, ISM, SJD, VR, JLW, KMW).

References

1. Ostrom QT, Gittleman H, Liao P, et al. CBTRUS Statistical Report: Primary brain and other central nervous system tumors diagnosed in the United States in 2010–2014. *Neuro-oncology*. 2017;19(suppl_5):v1–v88.
2. Johnson KJ, Cullen J, Barnholtz-Sloan JS, et al. Childhood brain tumor epidemiology: a brain tumor epidemiology consortium review. *Cancer Epidemiol Biomarkers Prev*. 2014;23(12):2716–2736.
3. Hagel C, Stemmer-Rachamimov AO, Bornemann A, et al. Clinical presentation, immunohistochemistry and electron microscopy indicate neurofibromatosis type 2-associated gliomas to be spinal ependymomas. *Neuropathology*. 2012;32(6):611–616.
4. Rubio MP, Correa KM, Ramesh V, et al. Analysis of the neurofibromatosis 2 gene in human ependymomas and astrocytomas. *Cancer Res*. 1994;54(1):45–47.
5. Ruggieri M, Praticò AD, Serra A, et al. Childhood neurofibromatosis type 2 (NF2) and related disorders: from bench to bedside and biologically targeted therapies. *Acta Otorhinolaryngol Ital*. 2016;36(5):345–367.
6. Torres CF, Korones DN, Pilcher W. Multiple ependymomas in a patient with Turcot's syndrome. *Med Pediatr Oncol*. 1997;28(1):59–61.
7. Mullins KJ, Rubio A, Myers SP, Korones DN, Pilcher WH. Malignant ependymomas in a patient with Turcot's syndrome: case report and management guidelines. *Surg Neurol*. 1998;49(3):290–294.
8. 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.
9. Gröbner SN, Worst BC, Weischenfeldt J, et al; ICGC PedBrain-Seq Project; ICGC MMML-Seq Project. The landscape of genomic alterations across childhood cancers. *Nature*. 2018;555(7696):321–327.
10. Leece R, Xu J, Ostrom QT, Chen Y, Kruchko C, Barnholtz-Sloan JS. Global incidence of malignant brain and other central nervous system tumors by histology, 2003-2007. *Neuro Oncol*. 2017;19(11):1553–1564.
11. Ostrom QT, Gittleman H, Truitt G, Boscia A, Kruchko C, Barnholtz-Sloan JS. CBTRUS statistical report: primary brain and other central nervous system tumors diagnosed in the united states in 2011–2015. *Neuro Oncol*. 2018;20(suppl_4):iv1–iv86.
12. U.S. Cancer Statistics Working Group. United States Cancer Statistics: 1999–2015 Incidence and Mortality Web-Based Report. http://www.cdc.gov/uscs. Accessed November 19, 2018.
13. Surveillance Epidemiology and End Results (SEER) Program. SEER*Stat Software Version 8.3.5.; 2018. www.seer.cancer.gov/seerstat. Accessed November 19, 2018.
14. Fay MP. Approximate confidence intervals for rate ratios from directly standardized rates with sparse data. *Commun Stat Theory Methods*. 1999;28(9):2141–2160.

15. Fay MP, Tiwari RC, Feuer EJ, Zou Z. Estimating average annual percent change for disease rates without assuming constant change. *Biometrics.* 2006;62(3):847–854.
16. Wiemels JL, Walsh KM, de Smith AJ, et al. GWAS in childhood acute lymphoblastic leukemia reveals novel genetic associations at chromosomes 17q12 and 8q24.21. *Nat Commun.* 2018;9(1):286.
17. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience.* 2015;4:7.
18. Purcell S, Chang C. PLINK 1.9. https://www.cog-genomics.org/plink2. Accessed December 2, 2018.
19. Das S, Forer L, Schönherr S, et al. Next-generation genotype imputation service and methods. *Nat Genet.* 2016;48(10):1284–1287.
20. McCarthy S, Das S, Kretzschmar W, et al; Haplotype Reference Consortium. A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet.* 2016;48(10):1279–1283.
21. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 2009;19(9):1655–1664.
22. Cann HM, de Toma C, Cazes L, et al. A human genome diversity cell line panel. *Science.* 2002;296(5566):261–262.
23. Raj A, Stephens M, Pritchard JK. fastSTRUCTURE: variational inference of population structure in large SNP data sets. *Genetics.* 2014;197(2):573–589.
24. Pasaniuc B, Sankararaman S, Kimmel G, Halperin E. Inference of locus-specific ancestry in closely related populations. *Bioinformatics.* 2009;25(12):i213–i221.
25. Shriner D, Adeyemo A, Rotimi CN. Joint ancestry and association testing in admixed individuals. *PLoS Comput Biol.* 2011;7(12):e1002325.
26. Li MX, Yeung JM, Cherny SS, Sham PC. Evaluating the effective numbers of independent tests and significant *P*-value thresholds in commercial genotyping arrays and public imputation reference datasets. *Hum Genet.* 2012;131(5):747–756.
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. Diskin SJ, Capasso M, Schnepp RW, et al. Common variation at 6q16 within HACE1 and LIN28B influences susceptibility to neuroblastoma. *Nat Genet.* 2012;44(10):1126–1130.
29. Delaneau O, Marchini J, Zagury JF. A linear complexity phasing method for thousands of genomes. *Nat Methods.* 2011;9(2):179–181.
30. Marchini J, Howie B, Myers S, McVean G, Donnelly P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat Genet.* 2007;39(7):906–913.
31. Howie B, Fuchsberger C, Stephens M, Marchini J, Abecasis GR. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat Genet.* 2012;44(8):955–959.
32. Pruim RJ, Welch RP, Sanna S, et al. LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics.* 2010;26(18):2336–2337.
33. Bryc K, Durand EY, Macpherson JM, Reich D, Mountain JL. The genetic ancestry of African Americans, Latinos, and European Americans across the United States. *Am J Hum Genet.* 2015;96(1):37–53.
34. Amezcua L, Beecham AH, Delgado SR, et al. Native ancestry is associated with optic neuritis and age of onset in hispanics with multiple sclerosis. *Ann Clin Transl Neurol.* 2018;5(11):1362–1371.
35. Villano JL, Parker CK, Dolecek TA. Descriptive epidemiology of ependymal tumours in the United States. *Br J Cancer.* 2013;108(11):2367–2371.
36. Ostrom QT, Cote DJ, Ascha M, Kruchko C, Barnholtz-Sloan JS. Adult glioma incidence and survival by race or ethnicity in the United States from 2000 to 2014. *JAMA Oncol.* 2018;4(9):1254–1262.
37. Abadin SS, Zoellner NL, Schaeffer M, Porcelli B, Gutmann DH, Johnson KJ. Racial/ethnic differences in pediatric brain tumor diagnoses in patients with neurofibromatosis type 1. *J Pediatr.* 2015;167(3):613–20.e1.
38. Magalhães TR, Casey JP, Conroy J, et al. HGDP and HapMap analysis by Ancestry Mapper reveals local and global population relationships. *PLoS One.* 2012;7(11):e49438.
39. López-Abente G, Pollán M, Ardanaz E, Errezola M. Geographical pattern of brain cancer incidence in the Navarre and Basque Country regions of Spain. *Occup Environ Med.* 2003;60(7):504–508.
40. Kopelman NM, Stone L, Wang C, et al. Genomic microsatellites identify shared Jewish ancestry intermediate between Middle Eastern and European populations. *BMC Genet.* 2009;10:80.
41. Behar DM, Yunusbayev B, Metspalu M, et al. The genome-wide structure of the Jewish people. *Nature.* 2010;466(7303):238–242.
42. Qin H, Zhu X. Power comparison of admixture mapping and direct association analysis in genome-wide association studies. *Genet Epidemiol.* 2012;36(3):235–243.
43. Ostrom QT, Egan KM, Nabors LB, et al. Glioma risk associated with extent of estimated European genetic ancestry in African Americans and Hispanics. *Int J Cancer.* 2020;146(3):739–748.