Glioma risk associated with extent of estimated European genetic ancestry in African-Americans and Hispanics

Quinn T. Ostrom¹, Kathleen M. Egan², L. Burt Nabor³, Travis Gerke², Reid C. Thompson⁴, Jeffrey J. Olson⁵, Renato LaRocca⁶, Sajeel Chowdhary⁷, Jeanette E. Eckel-Passow⁸, Georgina Armstrong¹, John K. Wiencke⁹, Jonine L. Bernstein¹⁰, Elizabeth B. Claus¹¹,¹², Dora Il'yasova¹³-¹⁵, Christoffer Johansen¹⁶, Daniel H. Lachance¹⁷, Rose K. Lai¹⁸, Ryan T. Merrell¹⁹, Sara H. Olson¹⁰, Siegal Sadetzki²⁰,²¹, Joellen M. Schildkraut²², Sanjay Shete²³, Richard S. Houlston²⁴, Robert B. Jenkins²⁵, Margaret R. Wrensch⁹, Beatrice Melin²⁶, Christopher I. Amos²⁷, Jason T. Huse²⁸, Jill S. Barnholtz-Sloan²⁹*, Melissa L. Bondy¹*  
* These authors jointly supervised this work and are joint senior authors

1. Department of Medicine, Section of Epidemiology and Population Sciences, Dan L. Duncan Comprehensive Cancer Center, Baylor College of Medicine, Houston, Texas  
2. Division of Population Sciences, H. Lee Moffitt Cancer Center & Research Institute, Tampa, Florida  
3. Neuro-oncology Program, University of Alabama at Birmingham, Birmingham, Alabama  
4. Department of Neurological Surgery, Vanderbilt University Medical Center, Nashville, Tennessee  
5. Department of Neurosurgery, Emory University School of Medicine, Atlanta, Georgia  
6. Department of Hematology-Oncology, Norton Cancer Institute, Louisville, Kentucky.  
7. Neuro-Oncology Program, Lynn Cancer Institute, Boca Raton, Florida  
8. Division of Biomedical Statistics and Informatics, Mayo Clinic College of Medicine, Rochester, Minnesota  
9. Department of Neurological Surgery, School of Medicine, University of California, San Francisco, San Francisco, California  
10. Department of Epidemiology and Biostatistics, Memorial Sloan Kettering Cancer Center, New York, New York  
11. School of Public Health, Yale University, New Haven, Connecticut  
12. Department of Neurosurgery, Brigham and Women’s Hospital, Boston, Massachusetts
13. Department of Epidemiology and Biostatistics, School of Public Health, Georgia State University, Atlanta, Georgia, USA
14. Cancer Control and Prevention Program, Department of Community and Family Medicine, Duke University Medical Center, Durham, North Carolina
15. Duke Cancer Institute, Duke University Medical Center, Durham, North Carolina
16. Oncology clinic, Finsen Center, Rigshospitalet and Survivorship Research Unit, The Danish Cancer Society Research Center, Copenhagen, Denmark
17. Department of Neurology, Mayo Clinic Comprehensive Cancer Center, Mayo Clinic, Rochester, Minnesota
18. Departments of Neurology and Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, California
19. Department of Neurology, NorthShore University HealthSystem, Evanston, Illinois
20. Cancer and Radiation Epidemiology Unit, Gertner Institute, Chaim Sheba Medical Center, Tel Hashomer, Israel
21. Department of Epidemiology and Preventive Medicine, School of Public Health, Sackler Faculty of Medicine, Tel-Aviv University, Tel-Aviv, Israel
22. Department of Public Health Sciences, University of Virginia School of Medicine, Charlottesville, Virginia
23. Department of Biostatistics, University of Texas MD Anderson Cancer Center, Houston, Texas
24. Division of Genetics and Epidemiology, The Institute of Cancer Research, Sutton, Surrey, UK
25. Department of Laboratory Medicine and Pathology, Mayo Clinic Comprehensive Cancer Center, Mayo Clinic, Rochester, Minnesota
26. Department of Radiation Sciences, Umeå University, Umeå, Sweden
27. Institute for Clinical and Translational Research, Dan L. Duncan Comprehensive Cancer Center, Baylor College of Medicine, Houston, Texas
28. Department of Pathology, University of Texas MD Anderson Cancer Center, Houston, Texas
29. Case Comprehensive Cancer Center, Case Western Reserve University School of Medicine, Cleveland, Ohio

**Short title:** Association of glioma with European ancestry

**Corresponding authors:**
Melissa Bondy, PhD; Baylor College of Medicine, One Baylor Plaza, Houston, Texas, 77030; phone: 713-798-2953; email: mbondy@bcm.edu
Jill S. Barnholtz-Sloan, PhD; Case Comprehensive Cancer Center, Case Western Reserve University School of Medicine, 2-526 Wolstein Research Building, 2103 Cornell Road, Cleveland, Ohio 44106; phone: 216-368-1506; email: jsb42@case.edu

**Keywords:** Glioma; genetic epidemiology; genetic ancestry; genome-wide association study

**Abbreviations used:** AFR: African super-population; AFR≥0.4: ≥40% African ancestry; AMR: American super-population; AMR≥0.15: ≥15% Native American ancestry; CEPH: Centre d'Etude du Polymorphism Humain; CEU: Utah Residents with Northern and Western European Ancestry; CHB: Han Chinese in Beijing China; EA: European ancestry; EAS: East Asian super-population; EUR: European super-population; GA: Global ancestry; GBM: Glioblastoma; GICC: Glioma International Case-Control Study; GliomaSE: Glioma Southeast Case-Control Study; GWAS: Genome-wide association study; JPT: Japanese in Tokyo Japan; LA: Local ancestry; LD: Linkage disequilibrium; MAF: Minor allele frequency; PEL: Peruvians from Lima Peru; RAF: Risk allele frequency; SNP: Single nucleotide polymorphism; US: United States; YRI: Yoruba in Ibadan Nigeria.

**Figures:** 4
Funding: This work was supported by grants from the National Institutes of Health (grants R01CA116174, R01CA207972, R01CA139020, R01CA52689, P50CA097257, P30CA008748, and P30CA125123). Additional support was provided by the McNair Medical Institute at Baylor College of Medicine (Houston, Texas) and the Population Sciences Biorepository at Baylor College of Medicine. The University of Alabama at Birmingham Comprehensive Cancer Center Neuro-oncology Research Acceleration Fund contributed to genotyping for the GliomaSE consortium. In Sweden work was additionally supported by Acta Oncologica through the Royal Swedish Academy of Science (BM salary) and The Swedish Research council and Swedish Cancer foundation. QTO is supported by a Research Training Grant from the Cancer Prevention and Research Institute of Texas (CPRIT; RP160097T).

Novelty and Impact (75/75 words):
Glioma is very rare in non-White populations, and genome-wide association studies in glioma to date have included exclusively European ancestry populations. In this study, we use African-American and Hispanic cases and controls from two large glioma case-control studies to assess association between European ancestry and glioma in non-White populations. This analysis identified an association between
glioma and two regions previously identified in EA populations, and four novel regions, suggesting regions to target in future studies.

**Category**: Cancer Epidemiology
ABSTRACT (250/250 WORDS)

Glioma incidence is highest in non-Hispanic Whites, and to date, glioma genome-wide association studies (GWAS) to date have only included European ancestry (EA) populations. African-Americans and Hispanics in the US have varying proportions of EA, African (AA) and Native American ancestries (NAA). It is unknown if identified GWAS loci or increased EA is associated with increased glioma risk. We assessed whether EA was associated with glioma in African-Americans and Hispanics. Data were obtained for 832 cases and 675 controls from the Glioma International Case-Control Study and GliomaSE Case-Control Study previously estimated to have <80% EA, or self-identify as non-White. We estimated global and local ancestry using fastStructure and RFMix, respectively, using 1,000 genomes project reference populations. Within groups with ≥40% AA (AFR≥0.4), and ≥15% NAA (AMR≥0.15), genome-wide association between local EA and glioma was evaluated using logistic regression conditioned on global EA for all gliomas. We identified two regions (7q21.11, p=6.36x10⁻⁴; 11p11.12, p=7.0x10⁻⁴) associated with increased EA, and one associated with decreased EA (20p12.13, p=0.0026) in AFR≥0.4. In addition, we identified a peak at rs1620291 (p=4.36x10⁻⁶) in 7q21.3. Among AMR≥0.15, we found an association between increased EA in one region (12q24.21, p=8.38x10⁻⁴), and decreased EA in two regions (8q24.21, p=0.0010; 20q13.33, p=6.36x10⁻⁴). No other significant associations were identified. This analysis identified an association between glioma and two regions previously identified in EA populations (8q24.21, 20q13.33), and four novel regions (7q21.11, 11p11.12, 12q24.21, 20p12.13). The identifications of novel association with EA suggests regions to target for future genetic association studies.
INTRODUCTION

Glioma is the most commonly occurring malignant brain tumor in the United States (US), with an average annual age-adjusted incidence of 6.0 per 100,000 from 2010-2014, though incidence varies significantly by sex, race, and age.\textsuperscript{1} Glioma incidence is highest in countries with majority European ancestry (EA) populations, including northern Europe, the US and Canada.\textsuperscript{2} Within the US, incidence of these tumors is highest among non-Hispanic Whites.\textsuperscript{1,3}

Though these tumors are rare, they cause significant morbidity and mortality. There are few confirmed risk factors, and the vast majority of cases occur in individuals with no family history.\textsuperscript{4,5} Previous genome-wide association studies (GWAS) in primarily European ancestry (≥80%) individuals have identified 25 genetic risk loci for glioma, which in total is estimated to account for ~30% of heritable risk, suggesting that there are both undiscovered environmental (which accounts for ~75% of overall risk variance) and genetic risk (accounting for ~70% of heritable risk).\textsuperscript{6,7} Due to the rarity of glioma overall, and the decreased incidence of glioma in populations other than non-Hispanic Whites, there have been limited analyses done to identify risk factors or genetic variants associated with glioma risk in non-European populations.

The majority of genetic association studies in glioma to date have been conducted in majority EA or East Asian-ancestry populations, where included individuals have a high proportion of estimated genetic ancestry that can be traced to one of these continental populations.\textsuperscript{8-13} Previous analyses have attempted to compare allele frequencies of previously identified risk loci within reference data sets by ancestry groups in order to account for differences in incidence, but these have failed to identify new risk variants in non-European ancestry populations.\textsuperscript{14} Several candidate SNP studies have been conducted in East Asian populations, which have found novel associations in \textit{XRCC1/3}, \textit{ZGPAT}, \textit{SLC2A4RG}, and \textit{SBTB46}, as well as validated associations previously discovered in European ancestry populations in \textit{EGFR} and \textit{RTEL1}.\textsuperscript{15}

\textsuperscript{16} Previous analyses of somatic features in astrocytoma by self-identified race have found increased
prevalence of TP53 mutations and decreased prevalence of EGFR amplification among African Americans as compared with non-Hispanic Whites, but these studies largely pre-date current molecular classification of these tumors, and are based on limited sample sizes.¹⁷,¹⁸

Candidate SNP studies in East Asians have identified novel risk variants, suggesting that genetic susceptibility to glioma may vary by ancestry. While nearly all individuals can trace portions of their genetic ancestry to multiple distinct populations, admixed populations with genetic ancestry that is the result of two or more previously isolated populations coming into contact and interbreeding. In the US, African Americans and Hispanics have continental ancestry from Africa, Europe, and the Native Americans. The overall proportion of EA varies substantially among individuals within these populations.¹⁹-²² Over decades and centuries, chromosomes become mosaics of the ancestral chromosomes from which they arose. Patterns of continental ancestry can be examined both globally (averaged continental ancestry across the genome) and locally (probable continental origin of specific segments of DNA). Multiple methods have been developed for identifying these mosaic segments, using reference populations with known continental ancestry. In addition finding ancestry-specific SNPs associated with specific diseases and other characteristics, ancestry-specific differences in linkage disequilibrium (LD) can also be used to identify the causal variant(s) within chromosomal regions identified by GWAS.²³

Glioma is more commonly reported in European ancestry and self-identified White Non-Hispanic populations, which may be due to enrichment for glioma risk alleles within European ancestry populations. As a result, the presence of increased proportions of global European ancestry in African American and Hispanic populations may be associated with increased glioma risk, and patterns of local European ancestry may be used to identify novel glioma risk loci. Here we attempted to assess whether variation in European ancestry was associated with glioma risk in populations with a combination of European, African and Native American ancestry.
MATERIALS AND METHODS

This study was approved by the institutional review board at University Hospitals Cleveland Medical Center, as the IRB of record for Case Western Reserve University School of Medicine where the data were secured and analyzed. All sites received Institutional Review Board or ethical board approval to conduct the study, and informed consent was obtained from all participants.

Study population

In this study, data were combined from two glioma GWAS: the Glioma International Case-Control Study (GICC) and the Glioma South-East Case-Control Study (GliomaSE), which have both been described in detail in previous publications. Only individuals 18 years or older at time of diagnosis or consent were included in all analyses. The GICC was a study conducted by the Genetic Epidemiology of Glioma International Consortium that recruited glioma cases and healthy controls from 14 centers across the US and Europe between 2010 and 2013. Controls were recruited using three approaches: seven sites recruited visitors accompanying non-brain tumor cancer patients, four sites recruited clinic-based controls at general medical clinics, and three sites used population-based controls. Race and ethnicity information was not available for all individuals from GICC, and as a result all individuals previously estimated to have <80% European ancestry using FastPop and excluded from Melin, et al. were used for these analyses. GliomaSE was a multi-center study that recruit glioma cases and controls from five centers across the southeastern US. Controls were recruited using two approaches: friends or family members of identified glioma cases, a through telephone listings who were frequency matched on age, gender, race and zip code. Cases and controls over 18 years old that self-identified as Hispanic or African American were included from GliomaSE. Cases and controls were not individually matched, but all analyses were adjusted for sex and age. Individual data were collected through patient interviews, and histologic classification was abstracted from pathology reports issued at recruiting institutions. After we completed quality control, these combined datasets included 832 cases and 675 controls (See Table 1 for additional study characteristics).
Genotyping and imputation

GICC cases and controls were genotyped on the Illumina Oncoarray, which was customized to include glioma-specific candidate SNPs and previous GWAS hits. GliomaSE cases and controls were genotyped on the Affymetrix UK Biobank Axiom array. Details of DNA collection and processing are available in previous publications. All datasets have previously undergone standard GWAS quality control using PLINK, and duplicate and related individuals within datasets have been excluded (as described in Melin et al.). Both datasets were imputed using Eagle 2 and Minimac3 as implemented on the Michigan Imputation Server (http://imputationserver.sph.umich.edu) using the 1,000 genomes phase 3 as a reference population. Imputed SNPs were filtered to those with \( r^2 \geq 0.7 \), and individual genotype probabilities \( \geq 0.7 \), after which genotype probabilities were converted to hard calls for further analysis. Principal components analysis was performed using the union set of the genotyped SNPs (Supplementary Figure 1), and all analysis were adjusted for both study and the first two principal components (which significantly differed between cases and controls) in order to adjust for differences in genotypes due to analysis platforms.

Ancestry estimation and statistical analysis

We estimated the global and local ancestry the using the following reference populations: Yoruba in Ibadan Nigeria (YRI, African super-population [AFR]), Peruvians from Lima Peru (PEL, American super-population [AMR]), Han Chinese in Beijing China (CHB, East Asian super-population [EAS]), Japanese in Tokyo Japan (JPT, EAS), and Utah Residents (CEPH) with Northern and Western European Ancestry (CEU, European super-population [EUR]). Principal components analysis was used to compare distribution of study samples to reference samples prior to global ancestry estimation to confirm that individuals were clustering with expected reference populations (Supplementary Figure 1). We estimated global ancestry using all 1,000 genomes AMR populations, and PEL was identified as being the most distinct from European and East Asian populations. Global ancestry was estimated using all
genotyped SNPs remaining after QC procedures with minor allele frequency ≥5% in any reference population using fastStructure, an efficient algorithm that approximates that of STRUCTURE for use with genome-wide SNP data (Supplementary Figure 2). Overall ancestry proportions were estimated for all four continental ancestry populations for each individual. RFMix v2.03 (http://github.com/slowkonirfmix) was used to identify local ancestry structure across all SNPs in the imputed set using the following continental ancestry reference populations: AFR (YRI), AMR (PEL), EAS (CHB, JPT), and EUR (CEU). RFMix uses a sliding window inferring local ancestry within each window by using a conditional random field parameterized by random forests. Estimates were generated under the assumption of 15 generations since admixture (as estimated by Zaitlen, et al., under the assumption of assortive mating by ancestry), two expectation-maximization iterations were performed to refine local ancestry estimates. RFMix outputs both most-likely ancestry calls as well as posterior probabilities for each ancestry. Statistical analyses were performed in R 3.5.0, and figures were generated using LocusZoom and the following R packages: ggplot2 and pophelper.

Global ancestry proportions were compared within self-identified race/ethnicity groups, phenotypes, and studies using t tests. Differences in local European ancestry (EA<sub>Local</sub>) between cases and controls were evaluated using logistic regression conditioned on global European ancestry (EA<sub>Global</sub>) for all glioma in individuals with ≥40% global African ancestry, and ≥15% global Native American ancestry. Both EA<sub>Local</sub> and EA<sub>Global</sub> were included in all analyses as continuous variables. Due to small sample size, we were unable to conduct histology-specific analyses for local ancestry. Logistic regression models were adjusted for study (GliomaSE versus GICC, when sets were combined for those with ≥40% global African ancestry only), sex, age at diagnosis, and the first two principal components estimated using a combined dataset. Associations were considered statistically significant at was p<1.67x10<sup>-4</sup> (Bonferroni correction for 300 tests) for individuals with ≥40% global African ancestry, and at p<2.17x10<sup>-4</sup> (Bonferroni correction for 230 test) for the ≥15% global Native American ancestry set. See Supplementary Note 1 for details of estimates of independent tests and power for these analyses.
For selected prioritized regions, unconditional logistic regression models in SNPTEST adjusted for age, and sex were used to generate per-allele odds ratios, 95% confidence intervals, and p values. For those with ≥40% global African ancestry only, GICC and GliomaSE estimates were combined using fixed-effects meta-analysis in META. Estimates for ≥80% global European ancestry individuals were obtained from Melin, et al. Single-SNP associations for 25 previously identified risk loci were considered statistically significant at p<0.002 level (Bonferroni correction for 25 tests). See Supplementary Note 2 for estimates of power for these analyses.

RESULTS

Overall, global and local ancestry were estimated for 1,507 individuals, including 832 cases and 675 controls (Table 1, see Supplementary Figure 3 for individual estimates of global ancestry by racial/ethnic groups and study group). After examining the distribution of global ancestry probabilities within each self-identified racial/ethnic group, 40% African ancestry and 15% Native American ancestry were selected as the cut-offs for further analysis to exclude individuals with the highest levels of European ancestry within each group (see Supplementary Figure 3). There were 373 individuals (193 cases and 180 controls) with ≥40% global African ancestry (AFR≥0.4, 244 from GICC, 129 from GliomaSE). Of these cases, 114 (59.1%) were glioblastoma (GBM) and 74 (40.9%) were non-GBM, while 5 (%) were other glioma histologies. There was a significant association between self-identification as black and belonging to the AFR≥0.4 set (p<2.2x10^{-16}; sensitivity=96.9% and specificity=97.5%). Within individuals self-identified as African American, estimated EA\textsubscript{Global} was slightly higher in glioma cases (20.0%) as compared to controls (17.4%) but the difference was not statistically significant (p=0.6102, Supplementary Figure 4A). EA\textsubscript{Global} varied by study, and was lower in individuals recruited as part of GliomaSE (GICC cases EA\textsubscript{Global}=20.6%, controls EA\textsubscript{Global}=18.1%, p=0.3581; GliomaSE cases EA\textsubscript{Global}=19.4%, controls EA\textsubscript{Global}=14.3%, p=0.9417, Supplementary Figure 4C). EA\textsubscript{Global} was also higher in GBM cases (20.7%, p=0.1046), and non-GBM cases (20.0%, p=0.2889), but these differences were not statistically significant (Supplementary Figure 5).
There were 425 individuals (232 cases and 190 controls) with ≥15% global Native American ancestry (AMR≥0.15, 396 from GICC, 29 from GliomaSE). Due to small sample size (21 cases and 8 controls), GliomaSE samples were excluded from further analysis of the AMR≥0.15 set, for a total of 211 cases and 182 controls. Of these cases, 97 (46.0%) were GBM, and 108 (51.2%) were non-GBM (e.g. astrocytic or oligodendroglial tumors), while 6 (2.8%) were other glioma histologies. Within GICC, Hispanic self-identification was significantly associated with the AMR≥0.15 set (p<2.2x10^-16; sensitivity=76.0% and specificity=96.6%). Within individuals identified as Hispanic in the GICC set, estimated EA_Global was higher in glioma cases (59.7%) as compared to controls (55.5%, p=0.0108, Supplementary Figure 6C). EA_Global was also non-significantly higher in GBM cases (58.3%, p=0.1836), and significantly higher in non-GBM cases (60.8%, p=0.0076) (Supplementary Figure 7).

Single-SNP associations were examined at the 25 glioma risk loci previously identified in European ancestry populations (Figure 2, Supplementary Table 1). In the AFR≥0.4 set, 0/25 SNPs were statistically significant at the p<0.002 level (Bonferroni correction for 25 tests), while 2/25 SNPs were nominally significant at the p<0.05 level: rs723527 (7p11.2, EGFR, p=0.0118, OR=0.66, 95%CI=0.48-0.91), and rs648044 (11q23.2, ZBTB16, p=0.0464, OR=0.67, 95%CI=0.45-0.99). Overall correlation between effect estimates (log odds ratio) across the 25 risk loci between the AFR≥0.4 set and the >80% European ancestry set was weak (adjusted R^2=0.426, Figure 2). The effect size at these SNPs in AFR≥0.4 was further from the null than the associations detected in Melin, et al. In the AMR≥0.15 set, 1/25 SNPs were statistically significant at the p<0.002 level (Bonferroni correction for 25 tests), while 4/25 SNPs were nominally significant at the p<0.05 level: rs10069690 (5p15.33, TERT, p=0.0056, OR=1.56, 95%CI=1.14-2.13), rs55705857 (8q24.21, CCDC26, p=0.0049, OR=2.39, 95%CI=1.30-4.37), rs78378222 (17p13.1, TP53, p=0.0130, OR=6.75, 95%CI=1.50-30.44), and rs6010620 (20q13.33, RTELI, p=4.83x10^-5, OR=2.02, 95%CI=1.44-2.83). Overall correlation between effect estimates (log odds ratio) across the 25 risk loci in the between the AMR≥0.15 set and the >80% European ancestry set was moderate (adjusted R^2=0.833,
Figure 2). The effect size at these SNPs in AMR$_{≥0.15}$ was further from the null than the associations detected in Melin, et al.$^7$

SNPs within 500kb of these 25 SNPs (~34,000 total SNPs with MAF>0.05 and INFO>0.7), as well as SNPs previously identified in East Asian populations, were examined within both sets to assess whether population-specific associations exist (Supplementary Table 2). None of these associations met the threshold for statistical significance (p<1.47x10$^{-6}$, Bonferroni correction for 34,000 tests) There were two SNPs identified as nominally significant in AMR$_{≥0.15}$ where association in Melin, et al. was null, including one SNP previously identified in East Asians (rs730437, EGFR$^{16}$, p=0.0063, OR=0.67, 95%CI=0.50-0.89; Melin, et al.: p=0.8600, OR=1.00, 95%CI=0.96-1.03). One other nominally significant association was identified in EGFR in this set (rs56129111, p=0.0010, OR=1.68, 95%CI=1.23-2.28; Melin, et al.: p=0.6638, OR=1.01, 95%CI=0.97-1.05). In AFR$_{≥0.4}$, one nominally significant association was identified in AKAP6 (rs733978; p=4.91x10$^{-4}$, OR=1.91, 95%CI=1.33-2.75).

Within the AFR$_{≥0.4}$ set, there was a nominally significant association between increased EA$_{Local}$ and glioma at 7q21.11 (p=6.36x10$^{-4}$), and 11p11.12 (p=7.0x10$^{-4}$), and a nominally significant negative association between EA$_{Local}$ and glioma at 20q12.13 (p=0.0026) (Figure 3). Single SNP associations were examined within these three regions (Supplementary Figure 8A-C), and a significant peak was identified at 7q21.3 with the strongest association at the C allele of rs1620291 (p=4.36x10$^{-6}$, OR=2.16, 95%CI=1.55-3.00, Figure 4B). In the Melin, et al. analysis this association was null (p=0.9151, OR=1.00, 95%CI=0.96-1.04). The allele frequency at this SNP is 0.34 in the 1,000 AFR population as compared to 0.65 in the EUR population (Figure 4C). Within AFR$_{≤0.4}$, the MAF in cases was close to the average of AFR and EUR (MAF=0.45), as compared to in controls where it was more similar to the AFR population (MAF=0.30). No other significant single SNP associations were identified.

Within the AMR$_{≤0.15}$ set, there was a nominally significant association between increased EA$_{Local}$ and glioma at 12q24.21 (p=8.38x10$^{-4}$), a nominally significant negative association between EA$_{Local}$ and glioma at 8q24.21 (p=0.0010) and 20q13.33 (p=6.36x10$^{-4}$) (Figure 3). Single SNP associations were
examined within these three regions, but no apparent peaks were identified (Supplementary Figure 8D-F).

DISCUSSION

This study represents the first GWAS to assess the relationship between European ancestry and glioma in admixed African-American and Hispanic populations in two multi-center case control studies. All previously conducted glioma GWAS have been conducted in predominantly European ancestry populations,\(^8\text{-}^{13}\) and non-European ancestry populations have been systematically excluded from GWAS in most complex diseases.\(^{42}\) Due to the rarity of glioma overall and the decreased incidence of glioma in African Americans as compared to persons of European ancestry there have been limited analyses done to identify genetic variants associated with glioma risk in populations with large proportions of non-European genetic ancestry.

In general, self-identified African American and Hispanic cases trended toward higher EA\(_\text{Global}\) as compared to controls, though many of these differences were not statistically significant. The mean level of African ancestry observed in African American cases in GICC and GliomaSE (78.7% in cases and 79.9% in controls) is similar to what was observed in a large-scale study conducted by 23andMe (73%).\(^{43}\) The mean level of Native American ancestry observed in Hispanic cases in GICC is (36.7% in cases and 37.5% in controls) is substantially higher than that observed in the same study of 23andMe data (18%).

This analysis found that mean African ancestry is highest and mean European ancestry is lowest among self-identified African Americans in the south east US. African -American cases and controls recruited for GliomaSE had lower levels of EA\(_\text{Global}\), which is consistent with the geographic location of these study sites.\(^{43}\) Similarly, the highest level of mean Native American ancestry among Hispanics was found in Texas and the southwest US, with relatively low levels of Native American ancestry among Hispanics in the eastern part of the US. The genotype quality and ancestry cutoffs used for these analyses were chosen to maximize sample size as well as overlap between the two datasets. There is no established standard for
defining African or Native American ancestry groups based on proportion of continental ancestries, and as a result the choice of boundaries used to define these groups may affect the results of the analysis.

No associations with the 25 previously identified GWAS hits reached statistical significance in either racial group, though some associations were nominally significant (Table 2). There were nominally significant associations observed with previously identified SNPs in \textit{ZBTB16} and \textit{EGFR} in the AFR$_{\geq0.4}$ set, and with \textit{TERT}, \textit{EGFR}, \textit{CCDC26}, \textit{TP53}, and \textit{RTEL1} in the AMR$_{\geq0.15}$ set. While other associations did not meet genome-wide significance, most associations were in a similar direction as those observed in Melin, et al.,$^7$ and it is possible that these associations may be significant with increased power or when analyses are stratified by histology. When the regions containing these SNPs were examined, there was an additional nominal association identified in \textit{AKAP6} in the AFR$_{\geq0.4}$ set, suggesting that there may be multiple population-specific SNPs tagging a causal variant or multiple causal variants increasing glioma risk. An \textit{EGFR} SNP previously identified in East Asians (rs730437) was nominally significant in the AMR$_{\geq0.15}$ set, but the direction of the association was the reverse of previously observed in the East Asian population. These results suggest that the SNPs previously identified by GWAS in other populations may not have identified the ‘true’ causal SNP, and that this SNP may tag different SNPs in different populations. Further fine-mapping of these loci in multi-ethnic populations may improve the resolution for detecting the causal SNP.$^{23}$ Another explanation is that there may be multiple causal SNPs within this region, and that these causal SNPs vary by ancestry group due to patterns of LD and allele frequencies.

This analysis identified a novel candidate association with increased EA$_{\text{local}}$ at 7q21 in the AFR$_{\geq0.4}$ set. Single SNP analyses identified a nominally significant association with rs1620291, an intergenic variant located within the antisense RNA \textit{AC002451.3} and upstream of pyruvate dehydrogenase kinase 4 [\textit{PDK4}] (7q21.3, upstream of see Figure 4A for genomic context). Multiple risk loci for breast cancer (rs17268829, rs111307654)$^{44}$, and prostate cancer (rs6465657)$^{45}$ have been identified within this region. This SNP lies within a previously identified 12 Mb structural variation hotspot on 7q,$^{46}$ in which insertions and deletions have been previously associated with genomic disorders. Local ancestry
analyses also identified nominally significant associations between \( EA_{\text{Local}} \) and glioma at 11p11.12 and 20p12.13.

Nominally significant associations were identified in the AMR\(\geq0.15 \) set between glioma and decreased \( EA_{\text{Local}} \) at 8q24.21, and 20q13.33, and increased \( EA_{\text{Local}} \) at 12q24.21. SNPs at 8q24.21 (rs55705857) and 20q13.33 (rs6010620) have both been previously associated with glioma in GWAS of majority European ancestry populations, and these previously identified SNPs were nominally significant in AMR\(\geq0.15 \) (rs55705857: \( p=0.0049, \text{OR}=2.39, \text{95\%CI}=1.30-4.37; \) rs6010620: \( p=4.83 \times 10^{-5}, \text{OR}=2.02, \text{95\%CI}=1.44-2.83 \)). The 8q24.21 region has been associated with multiple cancer types in GWAS in majority European ancestry populations. Prior GWAS have identified risk associations at 8q24.21 locus in East Asians, Latin Americans and African Americans for prostate and colon cancer.\(^{47-49}\)

This study has several limitations. While this sample does represent the largest dataset of genotyped non-European glioma cases, the small sample size limits the power of this analysis to detect significant associations between \( EA_{\text{Local}} \) and glioma. Due to the limited sample size, the analysis was limited only to a pooled assessment of glioma and not specific subtypes. The African-American and Hispanic sets had differing proportions of patients with GBM and non-GBM, which may limit the comparability of these two sets for loci with histology-specific associations. Due to limited sample size and the rarity of these cases, no validation set was available for this study. Increases in sample size are necessary in order to confirm the associations detected in this analysis, which due to the rarity of these cases necessitates additional multi-center collaborations.

Recruitment for GliomaSE and GICC in different regions of the US, which contributes to variation in \( EA_{\text{Global}} \) within race/ethnicity groups, particularly in regards to proportions of global African ancestry by study. This regional variation also likely results in heterogeneity in the specific populations contributing to Native American ancestry.\(^{50,51}\) Both global and local ancestry estimations are highly sensitive to the reference populations and settings used to generate these estimates. American reference populations are
all derived from admixed populations, and as a result ‘true’ proportion of Native American ancestry in those samples may be estimated incorrectly.

CONCLUSIONS

The results of this study suggest that increased European ancestry in admixed populations may be associated with increased risk of glioma. The identification of novel SNP associations within previously identified glioma risk regions may assist in fine mapping of these regions to identify causal variants. In order to accrue the larger sample sizes necessary for further discovery in this rare disease in minority populations, the development of further multi-institutional collaborations is necessary.

ACKNOWLEDGEMENTS

This work was previously presented at the Society for Neuro-Oncology 2017 Annual Meeting, the American Association for Cancer Research 2018 Annual Meeting, and the Brain Tumor Epidemiology Consortium 2018 Annual Meeting.

This work was supported by grants from the National Institutes of Health (grants R01CA116174, R01CA207972, R01CA139020, R01CA52689, P50CA097257, P30CA008748, and P30CA125123). Additional support was provided by the McNair Medical Institute at Baylor College of Medicine (Houston, Texas) and the Population Sciences Biorepository at Baylor College of Medicine. The University of Alabama at Birmingham Comprehensive Cancer Center Neuro-oncology Research Acceleration Fund contributed to genotyping for the GliomaSE consortium. In Sweden work was additionally supported by Acta Oncologica through the Royal Swedish Academy of Science (BM salary) and The Swedish Research council and Swedish Cancer foundation. QTO is supported by a Research Training Grant from the Cancer Prevention and Research Institute of Texas (CPRIT; RP160097T).
We are grateful to all the patients and individuals for their participation and we would also like to thank the clinicians and other hospital staff, cancer registries and study staff in respective centers who contributed to the blood sample and data collection.

REFERENCES


29. Purcell S, Chang C. PLINK 1.9.


FIGURE KEY

Figure 1. Schematic of data processing and imputation

Figure 2. Odds ratios, overall correlation, and –log_{10} p-values in non-European ancestry (EA) groups between estimates for Europeans (>=80% global European Ancestry) as compared to A) African-Americans (>=40% global African ancestry), and B) Hispanics (>=15% global Native American ancestry)

Figure 3. -log_{10} p-values for association between local European ancestry (EA) estimate and glioma in those with A) African-Americans (>=40% global African ancestry, GICC & GliomaSE), and B) Hispanics (>=15% global Native American ancestry, GICC only), and odds ratios and 95% confidence intervals by ancestry group associated with region of lowest p value as identified in C) African-Americans (>=40% global African ancestry, GICC & GliomaSE), and D) Hispanics (>=15% global Native American ancestry, GICC only)

Figure 4. A) Single SNP associations from 95Mb to 96Mb on chromosome 7 in AFR_{>=0.4} annotated with linkage disequilibrium in 1,000 genomes African ancestry super-population, B) Odds ratio and 95% confidence interval for rs1620291 by ancestry group, and C) alternate allele [C] frequencies for rs1620291 by 1000 genomes reference population and ancestry group