

ARTICLE

Trans-population Analysis of Genetic Mechanisms of Ethnic Disparities in Neuroblastoma Survival

Eric R. Gamazon, Navin Pinto, Anuar Konkashbaev, Hae Kyung Im, Sharon J. Diskin, Wendy B. London, John M. Maris, M. Eileen Dolan, Nancy J. Cox, Susan L. Cohn

Manuscript received July 3, 2012; revised October 17, 2012; accepted October 18, 2012.

Correspondence to: Nancy J. Cox, PhD, Section of Genetic Medicine, University of Chicago, 900 E 57th St, Rm 3220F, Chicago, IL 60637 (e-mail: ncox@medicine.bsd.uchicago.edu) and Susan L. Cohn, MD, Section of Pediatric Hematology/Oncology, University of Chicago, 900 E 57th St, Rm 5100, Chicago, IL 60637 (e-mail: scohn@peds.bsd.uchicago.edu).

Background Black patients with neuroblastoma have a higher prevalence of high-risk disease and worse outcome than white patients. We sought to investigate the relationship between genetic variation and the disparities in survival observed in neuroblastoma.

Methods The analytic cohort was composed of 2709 patients. Principal components were used to assign patients to genomic ethnic clusters for survival analyses. Locus-specific ancestry was calculated for use in association analysis. The shorter spans of linkage disequilibrium in African populations may facilitate the fine mapping of causal variants in regions previously implicated by genome-wide association studies conducted primarily in patients of European descent. Thus, we evaluated 13 single nucleotide polymorphisms known to be associated with susceptibility to high-risk neuroblastoma from genome-wide association studies and all variants with highly divergent allele frequencies in reference African and European populations near the known susceptibility loci. All statistical tests were two-sided.

Results African genomic ancestry was associated with high-risk neuroblastoma ($P = .007$) and lower event-free survival ($P = .04$, hazard ratio = 1.4, 95% confidence interval = 1.05 to 1.80). rs1033069 within *SPAG16* (sperm associated antigen 16) was determined to have higher risk allele frequency in the African reference population and statistically significant association with high-risk disease in patients of European and African ancestry ($P = 6.42 \times 10^{-5}$, false discovery rate < 0.0015) in the overall cohort. Multivariable analysis using an additive model demonstrated that the *SPAG16* single nucleotide polymorphism contributes to the observed ethnic disparities in high-risk disease and survival.

Conclusions Our study demonstrates that common genetic variation influences neuroblastoma phenotype and contributes to the ethnic disparities in survival observed and illustrates the value of trans-population mapping.

J Natl Cancer Inst;2013;105:302–309

Neuroblastoma is the most common extracranial solid malignancy in children and is notable for its clinical heterogeneity. Clinical features, including age at diagnosis and stage of disease, and genetic tumor markers are strongly predictive of outcome and used to define risk groups for treatment stratification (1). High-risk neuroblastoma is defined largely by older patients (aged >18 months) with metastatic disease, although all patients with amplification of the *MYCN* oncogene, regardless of stage, are considered high risk. Ethnic disparities in outcome are well described for many cancers (2–5). However, until recently, little was known about associations between race/ethnicity and survival in neuroblastoma. In 2011, Henderson and colleagues analyzed a cohort of 3539 neuroblastoma patients with known self-reported race and outcome and demonstrated for the first time that black children have an increased prevalence of high-risk disease and inferior survival compared with white children (6).

Multiple factors are known to contribute to racial health disparities, including differences in access to healthcare and other environmental and socioeconomic factors. However, because low-risk neuroblastomas only rarely progress to high-risk disease over time, we hypothesized that genetic variation is largely responsible for the ethnic disparities in tumor phenotype and outcome observed in neuroblastoma. Indeed, germline DNA variants associated with susceptibility to high- or low-risk neuroblastoma have been identified in genome-wide studies conducted primarily in children of European descent (7–12). To investigate the role African genomic ancestry plays in determining neuroblastoma phenotype and to identify genetic variants that statistically account for the ethnic disparities in outcome, we analyzed the genotypes of an ethnically admixed population of neuroblastoma patients (N = 2709) accrued through Children's

Patients and Methods

Patient Cohort

After institutional review board approval from participating institutions and informed consent were obtained, children diagnosed with neuroblastoma, ganglioneuroblastoma, or ganglioneuroma (maturing type) were enrolled in Children's Oncology Group ANBL00B1 (NCT00904241) between 2001 and 2009. Those patients with available genotype and outcome data formed the analytic cohort (Table 1). Methods to confirm the diagnosis, assignment of stage, analysis of tumor biology (ploidy, *MYCN* amplification and histology), and assignment of self-reported race have been previously described (6).

Genotyping

DNA samples from patients enrolled in Children's Oncology Group ANBL00B1 were genotyped using three Illumina platforms: 550v1, 550v3, and Human610-Quad, (Illumina, Inc., San Diego, CA) as previously described (7–9). Included in this analysis are those single nucleotide polymorphisms (SNPs) that were genotyped on all three platforms, had call rates greater than 95%, and had minor allele frequencies less than 5%.

Statistical Analysis

Quality Control and Data Processing. Supplementary Figure 1 (available online) is a diagram that illustrates the sample-based quality control pipeline, as previously described (13). Briefly, extensive quality control analyses were performed on the genotype data, including detection of sex incompatibilities, mis-specified relationships, and duplications.

Call rates were estimated by individual and by SNP to determine average heterozygosity (across all SNPs) for each sample and to evaluate genotype distributions for each SNP in relation to expected Hardy–Weinberg equilibrium (7–9,12).

Global Ancestry Quantification. Principal component analysis was conducted on the genotypes of 2919 individuals—2709 children with neuroblastoma plus 210 reference HapMap (14) samples from descendants of Northern Europeans (Utah residents with ancestry from Europe, CEU, $n = 60$), West Africans (Yoruba in Ibadan, Nigeria, YRI, $n = 60$) and East Asians (Han Chinese in Beijing, CHB, $n = 45$; Japanese in Tokyo, JPT, $n = 45$)—using EIGENSTRAT (15). The first principal component (PC1) separated patients with African ancestry from other patients. Principal component 2 (PC2) separated patients with Asian ancestry and patients with European ancestry. These leading principal components (PCs) provide the relative contribution of (geographically defined) reference populations.

Using the first 2 PCs, we applied a k -means clustering algorithm ($k = 5$) to classify patients into genomic ethnic clusters (Figure 1) (16).

In addition to principal component analysis, we used a model-based approach for ancestry quantification, as implemented in ADMIXTURE (17), assuming three founder populations. This approach simultaneously yields estimates of ancestry proportions and population allele frequencies in the samples. The latter enabled

us to calculate the extent of divergence between the estimated ancestral populations through the Fisher fixation index F_{ST} statistic. We calculated the Spearman correlation between PC1 and the proportion of African ancestry derived from ADMIXTURE to test the robustness of our ancestry estimation.

PC1, which separated the CEU and YRI populations, was tested for association with neuroblastoma risk group using logistic regression and with event-free survival using a Cox regression model. The proportional hazards assumption of the Cox regression was validated using *cox.zph* in the R survival package (18). Furthermore, Kaplan–Meier event-free survival analyses were performed on the genomic ethnic clusters derived from the k -means clustering algorithm, and curves were compared using a log-rank test. Time to event was calculated as the time from diagnosis until the first occurrence of neuroblastoma relapse, progression, or death. Patients without an event were censored at the time of last contact.

Genotype Association at Known High-Risk Neuroblastoma Susceptibility Loci With High Population Divergence

Contrasting association results in samples across major geographic regions at known susceptibility loci may allow us to distinguish causal variants and identify variants that account for the observed ethnic disparities. Although the challenges of conducting genetic association studies in African populations are well recognized, the shorter spans of linkage disequilibrium (LD) and the greater levels of haplotype diversity in African populations may facilitate the fine mapping of causal variants responsible for the genome-wide associations first identified in samples of European ancestry (19).

Thus, variants within 2 Mb of all the replicated susceptibility loci on chromosomes 2, 6, and 11 [identified in all previous genome-wide association studies conducted primarily in children of European descent (7–9)] that show high divergence between the ancestral populations (ie, European and African), as quantified by the Fisher fixation index ($F_{ST} > 0.25$), were prioritized for further analysis. These variants were tested through logistic regression analyses, with the proportion of African ancestry as a covariate, for association with high-risk neuroblastoma in a case-only study (high risk vs non-high risk). This association analysis was restricted to patients in genomic ethnic clusters 1, 2, and 3 (along the European–African ancestry axis; $n = 2368$ patients). Applying a Bonferroni adjustment for multiple testing, the level of statistical significance was set at P less than or equal to .0002. False discovery rate was computed using Storey's Q value package (20).

To test whether our SNP associations were confounded by population stratification, we conducted a subgroup association analysis using only patients with genomic European ancestry. All statistical tests were two-sided.

Results

Ancestry Quantification and Association With Risk Group and Survival

In the principal component analysis, the reference HapMap populations (14)—European (CEU), African (YRI), and Asians (JPT+CHB)—clustered tightly when the first two PCs were plotted on an ancestry map. The samples of European and African

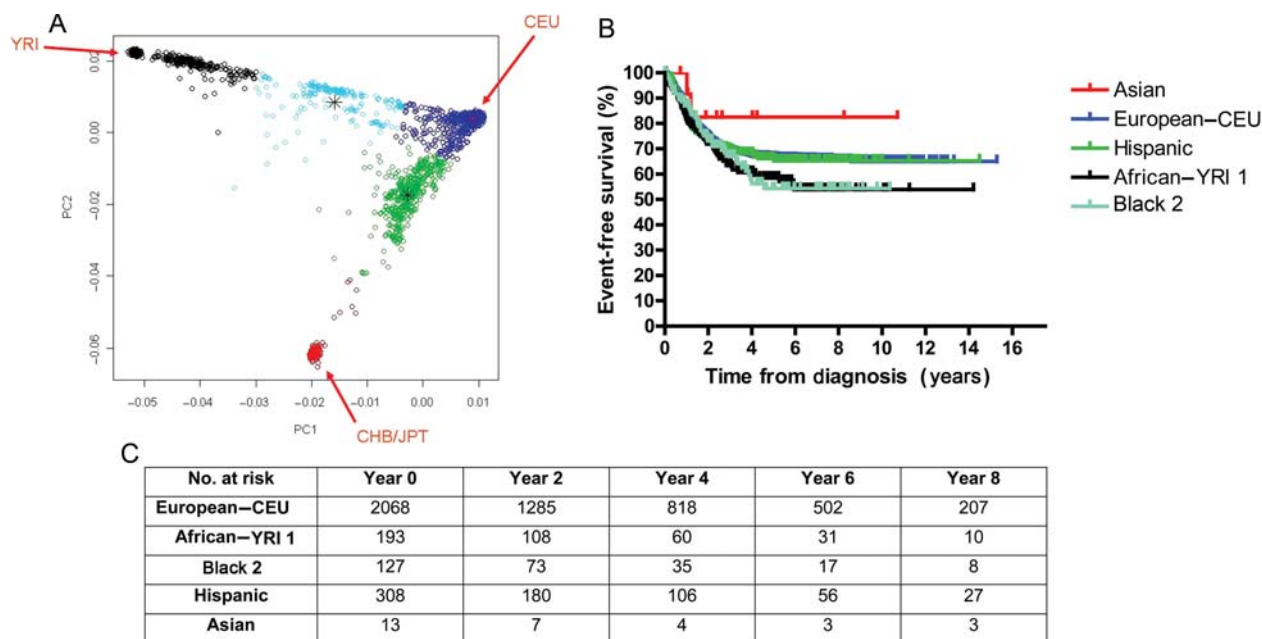


Figure 1. Principal components and event-free survival analyses. **A)** Principal components analysis and *k*-means clustering of the genotypes of 2709 children with neuroblastoma and 210 reference HapMap samples (red arrows). European reference is CEU, African reference is YRI, and Asian reference is CHB/JPT. **B)** Event-free survival by genomic cluster, with African cluster 1 and black cluster 2 corresponding to patients with the highest proportion of African ancestry and poorest event-free survival (two-sided log-rank $P = .03$). **C)** Number of patients at risk by genomic cluster.

descent in the study appeared along the “ancestry axis” between the CEU and YRI clusters. PC1 was associated with neuroblastoma risk group ($P = .007$), with African ancestry associated with high-risk disease. In addition, PC1 was also associated with outcome, with African ancestry associated with worse event-free survival ($P = .04$, hazard ratio = 1.4, 95% confidence interval = 1.05 to 1.80) in a Cox proportional hazard regression model. Adjustment for risk group in a Cox regression multivariable analysis abrogated the inferior outcome associated with PC1 ($P = .38$).

A *k*-means algorithm (16) was used to classify the patients into five genomic ethnic clusters for Kaplan–Meier survival analysis. As shown in Figure 1, event-free survival for clusters 1 (African–YRI genomic cluster) and 2 (African American genomic cluster), which consist of patients with the highest proportion of African ancestry, was statistically significantly worse than for the other ethnic clusters (log-rank $P = .03$).

In addition to principal component analysis, we used a model-based approach for ancestry estimation, as implemented in ADMIXTURE (17). Estimates of ancestry from both the ADMIXTURE and EIGENSTRAT methods showed a high degree of concordance (Spearman correlation = 0.997), which indicates the robustness of our estimates of ancestry. As expected from the near-perfect correlation, the estimate of the proportion of African ancestry from either method was associated with high-risk disease.

Population-Divergent Genetic Variables Associated With High-risk Neuroblastoma

We evaluated risk allele frequencies of 13 known susceptibility SNPs in *LINC00340* (7), *LMO1* (8), and *BARD1* (9) from previous genome-wide case–control studies using the HapMap CEU and YRI genotype data. Of the 13 candidate SNPs, six exhibited high

population differentiation ($F_{ST} > 0.25$). All but one (rs17487792, an intronic SNP within the *BARD1* locus) of these six SNPs showed higher risk allele frequencies in the reference African population vs the European cohort (Table 2).

To identify potential genetic-based mechanisms for the observed racial disparities in risk group classification and survival, we used a population genetics method applied to high-risk vs non-high-risk samples that cluster along the European–African ancestry axis (the European–CEU, African–YRI, and African American clusters from the *k*-means analysis; Figure 1), and prioritized variants near the known neuroblastoma susceptibility loci (7–9) with high population divergence ($F_{ST} > 0.25$). We tested each such SNP ($n = 245$ polymorphisms) near the known neuroblastoma susceptibility loci for association with high-risk phenotype using logistic regression with the proportion of African ancestry estimated by ADMIXTURE as a covariate. Table 3 lists the statistically significant associations from this analysis ($P < .0002$, Bonferroni method for multiple testing). The previously reported neuroblastoma susceptibility loci within *BARD1* (rs17487792) and on chromosome 6p22 (rs6939340, rs9295536 and rs4712653) were found to be statistically significantly associated with high-risk disease (7,9). Of note, the previously identified neuroblastoma susceptibility SNPs within *LMO1* (8) did not meet the statistical significance threshold in our analysis. Because the chromosome 6p22 SNPs do not lie in a protein-coding gene, they were excluded from further analyses and are the subject of ongoing investigation.

A novel *SPAG16* (sperm associated antigen 16) SNP (rs1033069) (LD $r^2 = 0.03$ with rs17487792 in CEU) was determined to be statistically significantly associated with high-risk neuroblastoma ($P = 6.42 \times 10^{-3}$, False Discovery Rate < 0.0015). Importantly, this SNP shows a substantially higher risk allele frequency in the

Table 1. Demographic and clinical characteristics of patients with neuroblastoma (N = 2709)*

Clinical characteristic	Entire cohort (N = 2709, 100%), No. (%)	African–YRI cluster 1† (n = 193, 7.1%), No. (%)	Black cluster 2† (n = 127, 4.7%), No. (%)	European–CEU cluster† (n = 2068, 76.3%), No. (%)	Hispanic cluster (n = 308, 11.4%), No. (%)	Asian–CHB/JPT cluster (n = 13, 0.5%), No. (%)	P‡
Age, months							
<18	1285 (47.4)	78 (40.4)	62 (48.8)	979 (47.3)	156 (50.6)	10 (76.9)	.045
>18	1424 (52.6)	115 (59.6)	65 (51.2)	1089 (52.7)	152 (49.4)	3 (23.1)	
Stage							
1, 2, 3, 4s	1459 (54.2)	94 (49.0)	67 (52.8)	1131 (55.1)	158 (51.6)	9 (69.2)	.30
4	1233 (45.8)	98 (51.0)	60 (47.2)	923 (44.9)	148 (48.4)	4 (30.8)	
Sex							
Female	1235 (45.6)	85 (44.0)	57 (44.9)	941 (45.5)	146 (47.4)	6 (46.2)	.96
Male	1474 (54.4)	108 (56.0)	70 (55.1)	1127 (54.5)	162 (52.6)	7 (53.8)	
MYCN status							
Not amplified	2067 (76.3)	139 (72.0)	105 (82.7)	1582 (76.5)	231 (75.0)	10 (76.9)	.56
Amplified	481 (17.8)	39 (20.2)	14 (11.0)	368 (17.8)	58 (18.8)	2 (15.4)	
Unknown	161 (5.9)	15 (7.8)	8 (6.3)	118 (5.7)	19 (6.2)	1 (7.7)	
Ploidy							
Hyperdiploid	1642 (60.6)	119 (61.7)	80 (63.0)	1252 (60.5)	185 (60.1)	6 (46.2)	.71
Diploid	805 (29.7)	51 (26.4)	32 (25.2)	623 (30.1)	93 (30.2)	6 (46.2)	
Unknown	262 (9.7)	23 (11.9)	15 (11.8)	193 (9.3)	30 (9.7)	1 (7.7)	
Histology							
Favorable	976 (36.0)	56 (29.0)	49 (38.6)	751 (36.3)	114 (37.0)	6 (46.2)	.45
Unfavorable	837 (30.9)	73 (37.8)	35 (27.6)	638 (30.9)	88 (28.6)	3 (23.1)	
Unknown	896 (33.1)	64 (33.2)	43 (33.9)	679 (32.8)	106 (34.4)	4 (30.8)	
Risk Group							
Low	960 (35.4)	52 (26.9)	41 (32.3)	753 (36.4)	108 (35.1)	6 (46.2)	.06
Intermediate	537 (19.8)	35 (18.1)	33 (26.0)	404 (19.5)	61 (19.8)	4 (30.8)	
High	1212 (44.7)	106 (54.9)	53 (41.7)	911 (44.1)	139 (45.1)	3 (23.1)	

* CEU = Utah residents with ancestry from Europe; CHB = Han Chinese in Beijing; JPT = Japanese in Tokyo; YRI = Yoruba in Ibadan, Nigeria.

† The first three clusters were included in the subgroup analysis of single nucleotide polymorphisms with high Fisher fixation index (between African and European ancestral groups) for association with risk group classification.

‡ P value is for overall comparison among all genomic clusters by two-sided Mantel–Haenszel χ^2 test.

Table 2. Ethnic variation in risk allele frequencies of single nucleotide polymorphisms (SNPs) previously implicated in the development of high-risk neuroblastoma*

SNP	Chromosome	F_{st} value†	Gene	Risk allele	CEU‡	YRI‡	CHB/JPT‡
rs17487792	2	0.27	<i>BARD1</i>	T	0.29	0.01	0.16
rs3768716	2	0.24	<i>BARD1</i>	G	0.24	0.02	0.18
rs6435862	2	—§	<i>BARD1</i>	—§	—§	—§	—§
rs6712055	2	0.05	<i>BARD1</i>	C	0.30	0.19	0.57
rs6715570	2	—§	<i>BARD1</i>	—§	—§	—§	—§
rs7587476	2	—§	<i>BARD1</i>	—§	—§	—§	—§
rs4712653	6	0.46	N/A	C	0.37	0.90	0.72
rs6939340	6	0.37	N/A	G	0.43	0.88	0.66
rs9295536	6	0.48	N/A	A	0.35	0.90	0.72
rs10840002	11	0.54	<i>LMO1</i>	A	0.30	0.90	0.50
rs110419	11	0.15	<i>LMO1</i>	A	0.50	0.82	0.61
rs204938	11	0.23	<i>LMO1</i>	C	0.47	0.81	0.19
rs4758051	11	0.44	<i>LMO1</i>	G	0.43	0.93	0.53

* CEU = Utah residents with ancestry from Europe; CHB = Han Chinese in Beijing; *FST* = Fisher fixation index; JPT = Japanese in Tokyo; YRI = Yoruba in Ibadan, Nigeria.

† The *FST* value for an SNP reflects allele frequency difference between the African and European ancestral groups.

‡ Denotes risk allele frequencies in corresponding reference populations.

§ Unless *FST* shows divergence, further information about the risk allele and its frequencies are not provided.

reference African vs European population. The risk allele frequency for the *SPAG16* SNP (rs1033069) is 0.34 in CEU and 0.76 in YRI. Further, a higher ancestral risk allele frequency in sub-Saharan Africa relative to Europe, Asia, and the Americas is observed in the geographic distribution of the rs1033069 allele frequencies from the Human Genome Diversity Panel populations (Figure 2)

(21–24). In contrast, the risk allele for the known susceptibility *BARD1* SNP (rs17487792), located 878 kb from the *SPAG16* SNP (rs1033069), is nearly absent in sub-Saharan Africa. Indeed, the risk allele frequency for this *BARD1* SNP in the HapMap YRI population is 0.008. Thus, this genetic variant does not explain the higher prevalence of high-risk disease in children with African vs

Table 3. Single nucleotide polymorphism (SNP) associations in genomic regions near known high-risk neuroblastoma susceptibility loci with high population divergence*

SNP	Chromosome	Gene	P†	Q	CEU‡	YRI‡	CHB/JPT‡
rs17487792	2	<i>BARD1</i>	5.89×10^{-8}	8.36×10^{-8}	0.29	0.01	0.16
rs1033067	2	<i>SPAG16</i>	5.52×10^{-5}	0.002	0.34	0.76	0.36
rs1033069	2	<i>SPAG16</i>	6.42×10^{-5}	0.002	0.34	0.76	0.36
rs6939340	6	N/A	1.88×10^{-5}	0.001	0.43	0.88	0.66
rs9295536	6	N/A	2.25×10^{-5}	0.001	0.35	0.90	0.72
rs4712653	6	N/A	5.56×10^{-5}	0.002	0.37	0.90	0.72

* CEU = Utah residents with ancestry from Europe; CHB = Han Chinese in Beijing; JPT = Japanese in Tokyo; N/A = Not Applicable; YRI = Yoruba in Ibadan, Nigeria.

† Each SNP was tested for association with high-risk phenotype using logistic regression with the proportion of African ancestry estimated by ADMIXTURE as a covariate. The likelihood ratio χ^2 statistic yielded a *P* value for each SNP. All tests were two-sided.

‡ Denotes risk allele frequencies in corresponding reference populations.

European ancestry. A nearby SNP in perfect LD with rs1033069 in both CEU and YRI—namely, rs1033067—showed a similar level of statistical significance.

Log-rank comparison of event-free survival by *SPAG16* SNP rs1033069 genotype (Figure 3) demonstrated that the number of risk alleles is statistically significantly associated with event-free survival (*P* = .007). This finding is consistent with the earlier survival analysis on the genomic ethnic clusters (Figure 1, B), which suggests that risk allele frequency differences at rs1033069 may be driving the observed disparities between the genomic ethnic clusters.

The *SPAG16* SNP was also associated with high-risk disease and survival in a subgroup analysis using only patients with genomic European ancestry or patients with self-reported European ancestry, indicating that the SNP association was not confounded by population stratification (Supplementary Table 1, available online). The association between the *SPAG16* SNP and high-risk disease/event-free survival in patients of European ancestry (*n* = 2068) becomes more statistically significant in the combined European and African sample set (*n* = 2368).

Functional Validation of rs1033069 and *SPAG16*

Because the risk allele for the *SPAG16* SNP (rs1033069) is common (minor allele frequency > 0.05) in groups with African ancestry, we hypothesized that additional polymorphisms in LD with the SNP may be identified using the 1000 Genomes Project dataset (see Supplementary Methods, available online) (25). In particular, we were interested in an expanded list of variants at this locus that exhibited large frequency differences between Africans and Europeans. Supplementary Table 2 (available online) summarizes the results of this analysis with annotation of effect on DNA regulatory motifs and chromatin state when applicable (26,27).

We also evaluated *SPAG16* mRNA abundance in lymphoblastoid cell lines derived from descendants of European ancestry (CEU) and African ancestry (YRI) (see Supplementary Methods, available online) and found higher levels of expression in the YRI cell lines (*P* = .04) (Supplementary Figure 2, available online).

Local Ancestry Estimates

The ADMIXTURE and EIGENSTRAT methods (see Patients and Methods) measure global ancestry across the genome and do

not estimate ancestry at a given locus. Indeed, individuals with similar global ancestry profiles or similar principal components may still have distinct local ancestry patterns (28). To perform local ancestry inference in populations formed by two-way admixture, we employed HAPMIX (see Supplementary Methods, available online) (29). We inferred the number of copies of each ancestry at each examined locus, which was then used in the association analysis as a covariate. Supplementary Table 3 (available online) shows the robustness of the associations (*P* < .0002, Bonferroni adjustment) to the use of locus-specific ancestry.

rs1033069 Genotype and Ethnic Disparities

Conditional analyses and a likelihood ratio approach (see Supplementary Methods, available online) show that the *SPAG16* SNP rs1033069 accounts for the ethnic disparities in phenotype (Supplementary Table 4, available online). Adjustment for rs1033069 substantially reduced the statistical significance of the racial disparities in the prevalence of high-risk phenotype (*P* = .12 after conditioning on rs1033069). In contrast, conditional analyses with the nearby *BARD1* SNP (rs17487792) did not statistically significantly alter the association between African ancestry and high-risk disease. Similarly, in a Cox proportional hazard regression model, the addition of the *SPAG16* SNP rs1033069 substantially reduced the statistical significance of the racial disparities associated with event-free survival (*P* = .11 after conditioning on rs1033069). Finally, the addition of rs1033069 to African genomic ancestry (PC1) fits the data statistically significantly better (*P* = 6.4×10^{-5}) in predicting high-risk disease, but the addition of PC1 to rs1033069 does not improve model fit (*P* = .12).

Discussion

Recently, racial disparities in outcome in neuroblastoma have been reported. A higher proportion of patients with high-risk disease were seen in self-reported black vs white populations, and in concordance with this disease presentation, black children had statistically significantly worse survival (6). We hypothesized that genetic variants in the genomic regions at known low-risk (11) or high-risk (7–9) neuroblastoma susceptibility loci (identified by previous genome-wide association study analyses performed primarily with patients of European descent) may statistically account for the observed racial disparities. To test this hypothesis, we evaluated

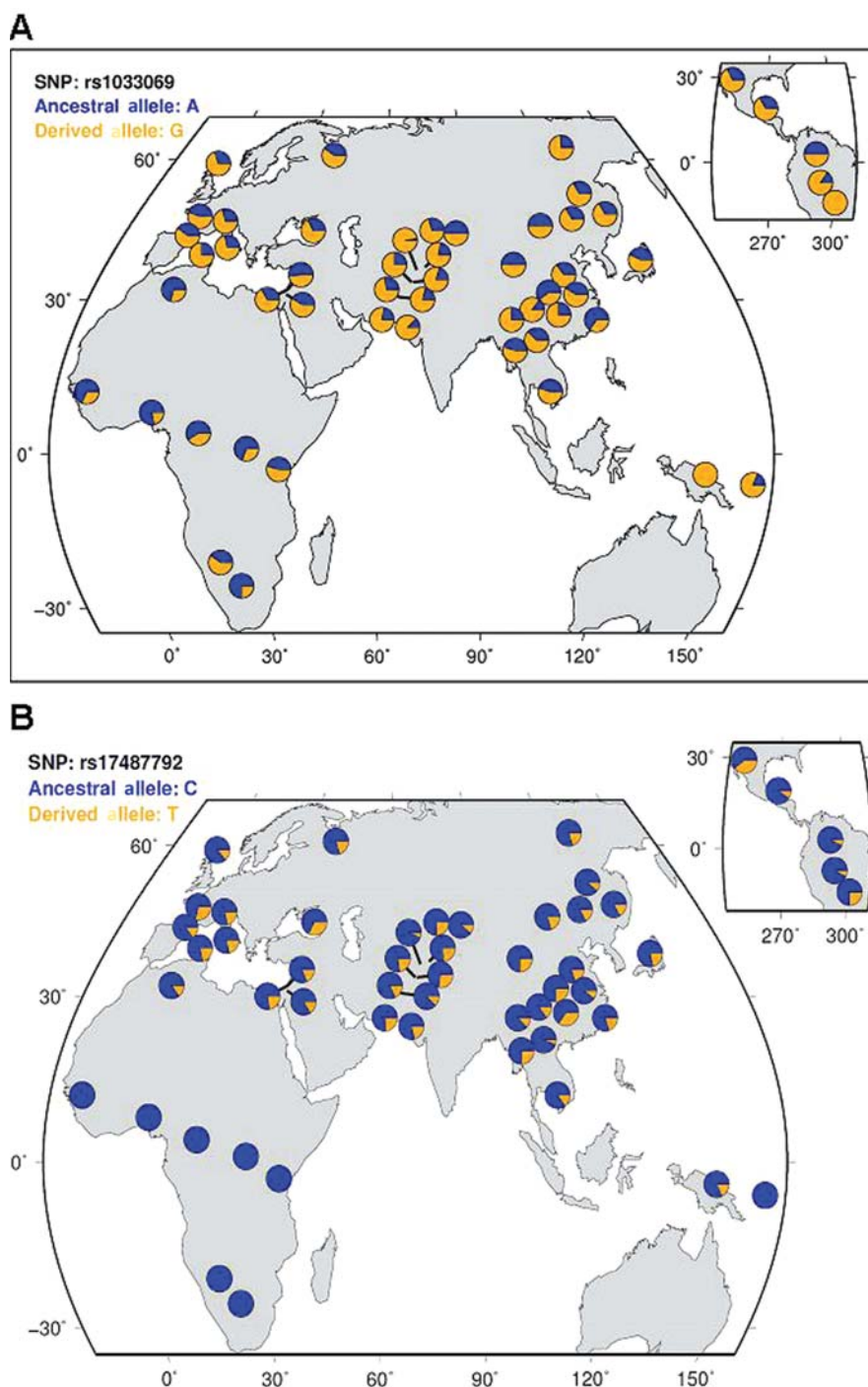
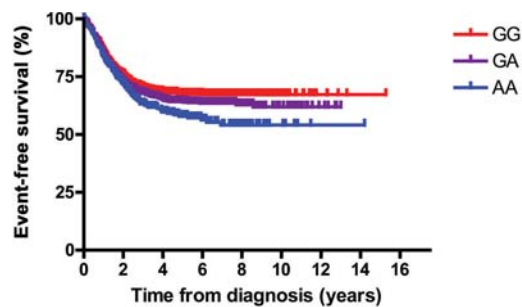


Figure 2. A comparison of the allele frequencies for a novel single nucleotide polymorphism (SNP) rs1033069 associated with high-risk neuroblastoma and the corresponding allele frequencies for the known nearby susceptibility SNP rs17487792, which was first identified in a genome-wide association study of patients of European descent. Note that a higher ancestral risk allele frequency in sub-Saharan Africa relative to Europe, Asia, and the Americas is observed in the geographic

distribution of the rs1033069 allele frequencies from the Human Genome Diversity Panel populations. **A)** Geographic distribution of allele frequencies for the novel disease-associated SNP rs1033069 at locus 2q34 where the ancestral allele (the A allele) is the risk allele. **B)** Geographic distribution of allele frequencies for the known susceptibility SNP rs17487792 at locus 2q35 where the derived allele (the T allele) is the risk allele.

the genotypes in an admixed population of children with neuroblastoma. The shorter spans of LD in African populations should facilitate the fine mapping of causal loci in previously implicated regions. We found that African genomic ancestry was statistically significantly associated with high-risk neuroblastoma phenotype and lower event-free survival. Multivariable analysis of the

association between African ancestry and event-free survival after accounting for diagnostic risk group showed that event-free survival failed to achieve statistical significance, suggesting that ethnic disparities in survival are driven by ethnic disparities in risk group stratification at diagnosis. We evaluated the variants near the known neuroblastoma susceptibility SNPs with high population



No. at risk	Year 0	Year 2	Year 4	Year 6	Year 8
GG	965	585	378	227	93
GA	1065	611	384	232	86
AA	338	194	113	63	27

Figure 3. Log-rank comparison of event-free survival by *SPAG16* (sperm associated antigen 16) single nucleotide polymorphism (SNP) rs1033069 genotype. Note that the number of risk alleles is statistically significantly associated with event-free survival (two-sided log-rank $P = .007$) with a higher number of the risk alleles for the SNP associated with worse outcome.

differentiation between the reference African and European populations. Five SNPs (three in *LMO1*, and two in *LINC00340*) known to be associated with the susceptibility to high-risk neuroblastoma were found to have high population divergence, with higher risk allele frequencies in the African cohort. We expanded our association analysis using a novel case-only design and were able to replicate previous high-risk susceptibility variants within *BARD1* and *LINC00340* and find additional variants within *SPAG16* associated with high-risk neuroblastoma. In conditional analyses, the *SPAG16* SNP, and not the nearby *BARD1* SNP, abrogated the racial disparities in risk-group and survival, which indicates that this variant may contribute to the higher prevalence of high-risk disease and inferior survival observed in children with African ancestry.

Neuroblastoma offers a unique opportunity to dissect genetic and environmental factors involved in racial disparities in risk and survival. In the United States, racial and ethnic differences in socioeconomic standing contribute to poorer access to care, delays in diagnosis, and poorer overall health outcomes (30). However, for children with neuroblastoma, diagnostic delays have limited influence on the tumor phenotype and ultimate risk-group classification. Observation studies have shown that it is unlikely for low-risk, favorable-biology tumors to progress to metastatic, high-risk tumors over time (31). Furthermore, *MYCN* status in neuroblastoma tumors, a powerful genetic prognostic factor, does not change over time (32,33). In addition, medication compliance problems are reduced in neuroblastoma because the majority of treatment (surgery, chemotherapy, radiation, immunotherapy) is delivered in the inpatient setting. Taken together, these observations suggest that socioeconomic and environmental factors are likely to play a more limited role in contributing to racial disparities in neuroblastoma than in other cancers.

In contrast to previous case-control genome-wide association studies, our study was a case-only design with the goal of localizing risk variants relevant to the development of high-risk neuroblastoma. Our analysis recapitulated several findings from earlier studies of neuroblastoma susceptibility, including SNPs within *BARD1* and on chromosome 6p22, but also identified novel high-risk neuroblastoma-associated SNPs at these loci in both the combined population and in samples of European descent.

These novel results implicate variants within the ubiquitously expressed *SPAG16* (34) as important in the acquisition of high-risk neuroblastoma. This study highlights the importance of expanding genetic association studies to non-European populations to enable identification of novel disease variants.

One limitation of our study is a lack of a replication cohort. Given the rarity of neuroblastoma and the relatively small number of patients of African ancestry genotyped in this study, we felt that inclusion of all patients was warranted to maximize power to detect associations.

Our findings emphasize the key role genomic variation may play in predicting outcome in children with neuroblastoma. Efforts to understand the underlying biology of these variants associated with risk group classification and survival are underway and may ultimately lead to the identification of more effective treatment strategies.

References

1. Maris JM, Hogarty MD, Bagatell R, et al. Neuroblastoma. *Lancet*. 2007;369(9579):2106–2120.
2. Ademuyiwa FO, Edge SB, Erwin DO, et al. Breast cancer racial disparities: unanswered questions. *Cancer Res*. 2011;71(3):640–644.
3. Allard JE, Maxwell GL. Race disparities between black and white women in the incidence, treatment, and prognosis of endometrial cancer. *Cancer Control*. 2009;16(1):53–56.
4. Downs LS, Smith JS, Scarinci I, et al. The disparity of cervical cancer in diverse populations. *Gynecol Oncol*. 2008;109(2 Suppl):S22–S30.
5. Goggins WB, Lo FF. Racial and ethnic disparities in survival of US children with acute lymphoblastic leukemia: evidence from the SEER database 1988–2008. *Cancer Causes Control*. 2012;23(5):737–743.
6. Henderson TO, Bhatia S, Pinto N, et al. Racial and ethnic disparities in risk and survival in children with neuroblastoma: a Children's Oncology Group study. *J Clin Oncol*. 2011;29(1):76–82.
7. Maris JM, Mosse YP, Bradfield JP, et al. Chromosome 6p22 locus associated with clinically aggressive neuroblastoma. *N Engl J Med*. 2008;358(24):2585–2593.
8. Wang K, Diskin SJ, Zhang H, et al. Integrative genomics identifies *LMO1* as a neuroblastoma oncogene. *Nature*. 2011;469(7329):216–220.
9. Capasso M, Devoto M, Hou C, et al. Common variations in *BARD1* influence susceptibility to high-risk neuroblastoma. *Nat Genet*. 2009;41(6):718–723.
10. Diskin SJ, Hou C, Glessner JT, et al. Copy number variation at 1q21.1 associated with neuroblastoma. *Nature*. 2009;459(7249):987–991.

11. Nguyen le B, Diskin SJ, Capasso M, et al. Phenotype restricted genome-wide association study using a gene-centric approach identifies three low-risk neuroblastoma susceptibility Loci. *PLoS Genet.* 2011;7(3):e1002026.
12. Latorre V, Diskin SJ, Diamond MA, et al. Replication of neuroblastoma SNP association at the BARD1 Locus in African-Americans. *Cancer Epidemiol Biomarkers Prev.* 2012;21(4):658–663.
13. Pluzhnikov A, Below JE, Konkashbaev A, et al. Spoiling the whole bunch: quality control aimed at preserving the integrity of high-throughput genotyping. *Am J Hum Genet.* 2010;87(1):123–128.
14. The International HapMap Consortium. The International HapMap Project. *Nature.* 2003;426(6968):789–796.
15. Price AL, Patterson NJ, Plenge RM, et al. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet.* 2006;38(8):904–909.
16. Ding C, He X. K-means clustering via principal component analysis. In: 21st International Conference on Machine Learning; July 4–8, 2004; Banff, Canada.
17. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 2009;19(9):1655–1664.
18. Grambsch PM, Therneau TM. Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika.* 1994;81(3):515–526.
19. Teo YY, Small KS, Kwiatkowski DP. Methodological challenges of genome-wide association analysis in Africa. *Nat Rev Genet.* 2010;11(2):149–160.
20. Storey JD, Tibshirani R. Statistical significance for genomewide studies. *Proc Natl Acad Sci U S A.* 2003;100(16):9440–9445.
21. Coop G, Pickrell JK, Novembre J, et al. The role of geography in human adaptation. *PLoS Genet.* 2009;5(6):e1000500.
22. Pickrell JK, Coop G, Novembre J, et al. Signals of recent positive selection in a worldwide sample of human populations. *Genome Res.* 2009;19(5):826–837.
23. Pritchard JK, Pickrell JK, Coop G. The genetics of human adaptation: hard sweeps, soft sweeps, and polygenic adaptation. *Curr Biol.* 2010;20(4):R208–R215.
24. Cann HM, de Toma C, Cazes L, et al. A human genome diversity cell line panel. *Science.* 2002;296(5566):261–262.
25. The 1000 Genomes Project Consortium. A map of human genome variation from population-scale sequencing. *Nature.* 2010;467(7319):1061–1073.
26. Ernst J, Kheradpour P, Mikkelsen TS, et al. Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature.* 2011;473(7345):43–49.
27. Ward LD, Kellis M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res.* 2012;40(database issue):D930–D934.
28. Wheeler HE, Gorsic LK, Welsh M, et al. Genome-wide local ancestry approach identifies genes and variants associated with chemotherapeutic susceptibility in African Americans. *PLoS One.* 2011;6(7):e21920.
29. Price AL, Tandon A, Patterson N, et al. Sensitive detection of chromosomal segments of distinct ancestry in admixed populations. *PLoS Genet.* 2009;5(6):e1000519.
30. Smedley BD, Stith AY, Nelson AR, eds. *Unequal Treatment: Confronting Racial and Ethnic Disparities in Health Care.* Washington, DC: National Academies Press; 2003.
31. Tanaka M, Kigasawa H, Kato K, et al. A prospective study of a long-term follow-up of an observation program for neuroblastoma detected by mass screening. *Pediatr Blood Cancer.* 2010;54(4):573–578.
32. Berthold F, Hero B, Breu H, et al. The recurrence patterns of stages I, II and III neuroblastoma: experience with 77 relapsing patients. *Ann Oncol.* 1996;7(2):183–187.
33. Brodeur GM, Hayes FA, Green AA, et al. Consistent N-myc copy number in simultaneous or consecutive neuroblastoma samples from sixty individual patients. *Cancer Res.* 1987;47(16):4248–4253.
34. Auer J, Senechal H, De Almeida M. Sperm-associated and circulating IgA and IgG classes of antibodies recognise different antigens on the human sperm plasma membrane. *J Reprod Immunol.* 1997;34(2):121–136.

Funding

This research was funded by the National Institutes of Health (U01 GM61393 [to MED and NJC], R01 MH090937 [to MED], U01 HG005773 [to NJC], R01 MH090937 [to NJC], and R01 CA078545 [to JMM]; Alex's Lemonade Stand (to SLC); Children's Neuroblastoma Cancer Foundation (to SLC); Elise Anderson Fund (to SLC); Neuroblastoma Children's Cancer Society (to SLC); Little Heroes Cancer Research Foundation (to SLC); St. Baldrick's Foundation (to NP); and Cancer Research Foundation (to NP).

Notes

E. R. Gamazon and N. Pinto contributed equally to this work.

Authors had full responsibility for the design of the study, the collection of the data, the analysis and interpretation of the data, the decision to submit the manuscript for publication, and the writing of the manuscript.

The authors have no conflict of interest to declare with respect to this manuscript.

Affiliations of authors: Department of Medicine, Section of Genetic Medicine (ERG, AK, NJC), Department of Pediatrics, Section of Pediatric Hematology/Oncology (NP, SLC), Department of Health Studies (HKI), and Department of Medicine, Section of Hematology/Oncology (ED), University of Chicago, Chicago, IL; Division of Oncology and Center for Childhood Cancer Research, Children's Hospital of Philadelphia and Department of Pediatrics, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA (SJD, JMM); Boston Children's Hospital / Dana Farber Harvard Cancer Center, Harvard University, Boston, MA (WBL); Children's Oncology Group Statistics and Data Center, Arcadia, CA (RBL).