Chromosome 6p22 Locus Associated with Clinically Aggressive Neuroblastoma


From the Division of Oncology and the Center for Childhood Cancer Research (J.M.M., Y.P.M., C.H., E.F.A., S.J.D., M.L., C.W., K.A.C., E.F.R.), the Center for Applied Genomics (J.P.B., J.T.G., C.K., E.C.F., T.C., A.W.E., S.F.A.G., H.H.), and the Division of Genetics (M.C., M.D., S.F.A.G., H.H.), Children’s Hospital of Philadelphia; and the Department of Pediatrics (J.M.M., Y.P.M., C.H., E.F.A., S.J.D., M.L., C.W., K.A.C., E.F.R.), the Abramson Family Cancer Research Institute, (J.M.M.), the Department of Biostatistics and Epidemiology (S.M., M.D., H.L.), and the Department of Pediatrics (M.C., M.D., S.F.A.G., H.H.), University of Pennsylvania School of Medicine — all in Philadelphia; the Section of Cancer Genetics, Institute of Cancer Research, Sutton, Surrey, United Kingdom (R.H.S., N.R.); the Division of Hematology–Oncology and the Saban Research Institute, Children’s Hospital Los Angeles, and the Keck School of Medicine, University of Southern California — both in Los Angeles (S.A., R.C.S.); Federico II University, Centro di Ingegneria Genetica Biotecnologie Avanzate, Naples, Italy (M.C.); the Division of Reproductive and Child Health and CRUK Institute for Cancer Studies, University of Birmingham, Birmingham, United Kingdom (C.M.); the Department of Statistics, University of Florida, and the Children’s Oncology Group — both in Gainesville (W.B.L.); and the Department of Experimental Medicine, La Sapienza University of Rome, Rome (M.D.). Address reprint requests to Dr. Maris at the Children’s Hospital of Philadelphia, 907A, 3516 Civic Center Blvd., Philadelphia, PA 19104-4318, or at maris@chop.edu, or to Dr. Hakonarson at the Children’s Hospital of Philadelphia, ARC12, 3516 Civic Center Blvd., Philadelphia, PA 19104-4318, or at hakonarson@chop.edu.

Dr. Mosse and Mr. Bradfield contributed equally to this article.

This article (10.1056/NEJMoa0708698) was published at www.nejm.org on May 7, 2008.
Copyright © 2008 Massachusetts Medical Society.

ABSTRACT

BACKGROUND
Neuroblastoma is a malignant condition of the developing sympathetic nervous system that most commonly affects young children and is often lethal. Its cause is not known.

METHODS
We performed a genomewide association study by first genotyping blood DNA samples from 1032 patients with neuroblastoma and 2043 control subjects of European descent using the Illumina HumanHap550 BeadChip. Samples from three independent groups of patients with neuroblastoma (a total of 720 patients) and 2128 control subjects were then genotyped to replicate significant associations.

RESULTS
We observed a significant association between neuroblastoma and the common minor alleles of three consecutive single-nucleotide polymorphisms (SNPs) at chromosome band 6p22 and containing the predicted genes FLJ22536 and FLJ44180 (P = 1.71 × 10^-9 to 7.01 × 10^-10; allelic odds ratio, 1.39 to 1.40). Homozygosity for the at-risk G allele of the most significantly associated SNP, rs6939340, resulted in an increased likelihood of the development of neuroblastoma (odds ratio, 1.97; 95% confidence interval, 1.58 to 2.45). Subsequent genotyping of the three 6p22 SNPs in three independent case series confirmed our observation of an association (P = 9.33 × 10^-5 at rs6939340 for joint analysis). Patients with neuroblastoma who were homozygous for the risk alleles at 6p22 were more likely to have metastatic (stage 4) disease (P = 0.02), amplification of the MYCN oncogene in the tumor cells (P = 0.006), and disease relapse (P = 0.01).

CONCLUSIONS
A common genetic variation at chromosome band 6p22 is associated with susceptibility to neuroblastoma.
Despite marked improvements in the cure rates for many childhood cancers, neuroblastoma remains an important clinical problem, accounting for 15% of the deaths attributable to malignant conditions in children. It is the most common solid cancer of early childhood, and approximately half of all patients with neuroblastoma present with widely disseminated disease that is often refractory to intensive chemoradiotherapy. Cure rates among these high-risk patients remain less than 40%, despite dramatic increases in the intensity of chemoradiotherapy, and survivors often have serious lifelong coexisting conditions.

Somatically acquired genomic aberrations in neuroblastoma are of fundamental importance for predicting the tumor phenotype in patients. Tumors with amplification of the MYCN oncogene or deletions of the chromosome arms 1p, 11q, or both, typically are metastatic at diagnosis and resistant to therapy. Conversely, tumors showing no structural chromosomal changes but hyperdiploidy due to whole-chromosome gains are more easily cured and may even spontaneously regress.

Neuroblastoma is an embryonal cancer; it is thought to arise from partially committed primordial cells during fetal or early childhood development. Despite the wealth of knowledge about somatically acquired genomic aberrations that correlate with tumor phenotype, little is known about the events that predispose to the development of neuroblastoma. Epidemiologic studies have not identified a common environmental exposure that influences susceptibility to neuroblastoma, and genetic studies of hereditary disease have been hampered by the rarity of the condition and the small size of pedigrees due to the lethality of neuroblastoma in early childhood. A family history of neuroblastoma is obtained in only about 1% of patients, and studies of such families suggest locus heterogeneity; no commonly mutated gene has been identified. We therefore hypothesized that neuroblastomas arise from relatively common DNA variations that predispose to an increased risk of neuroblastic malignant transformation.

**METHODS**

**SUBJECTS**

For genomewide genotyping, case subjects were defined as children with a diagnosis of neuroblastoma or ganglioneuroblastoma and registered through the Children's Oncology Group. The blood samples from the patients with neuroblastoma were identified through the neuroblastoma biorepository of the Children's Oncology Group for specimen collection at the time of diagnosis. The majority of specimens were annotated with clinical and genomic information that included the patient's age at diagnosis, site of tumor origin, disease stage according to the International Neuroblastoma Staging System, International Neuroblastoma Pathology Classification, MYCN oncogene copy number, DNA index (i.e., ploidy status), registration in a clinical trial or clinical trials, event-free and overall survival, second tumors, and any associated conditions (e.g., congenital abnormalities).

The eligibility criterion for genomewide genotyping was the availability of 1.5 μg of DNA of high quality from a tumor-free source such as peripheral blood or bone marrow mononuclear cells that were uninvolved with a tumor. Because neuroblastoma in the United States most often occurs in persons of European descent, we limited our initial analyses to the DNA samples from the blood of such persons to minimize genetic variability. We randomly divided subjects with samples dedicated to genomewide testing into a group of 1251 case subjects for the discovery phase and a group of 409 case subjects for an initial replication phase.

Control subjects were recruited from the Philadelphia region through the Children's Hospital of Philadelphia Health Care Network, including four primary care clinics and several group practices and outpatient practices that included well-child visits. Eligibility criteria for control subjects were European ancestry as determined by self-report or parental report, availability of 1.5 μg of high-quality DNA from peripheral-blood mononuclear cells, and no serious underlying medical disorder, including cancer. The median age of the control subjects at the time of sample collection was 10.0 years. We genotyped single-nucleotide polymorphisms (SNPs) across the genomes of 3414 subjects to the discovery phase and 1178 to the initial replication phase.

Two additional case series were used for the purpose of replication. First, 252 randomly selected and unrelated patients with neuroblastoma or ganglioneuroblastoma were recruited from pediatric oncology centers in the United Kingdom.
either as part of a group of patients who had received a diagnosis of neuroblastoma at the Birmingham Children’s Hospital since 1992, through the Factors Associated with Childhood Tumours study, or from the Tumour and Leukaemia Bank of the Childhood Cancer and Leukaemia Group. A total of 788 samples from control subjects were obtained from the 1958 Birth Cohort collection, an ongoing follow-up study of all persons born in Britain during 1 week in 1958, including a recent biomedical assessment during the period from 2002 through 2004; during this assessment, blood samples and written informed consent were obtained for the creation of a genetic resource (the National Child Developmental Study, www.cls.ioe.ac.uk/). All case subjects and control subjects were from the United Kingdom, and subjects known to be of non-European ancestry were excluded. The final replication group of 59 unrelated persons with high-risk neuroblastoma were recruited from the U.S.-based Children’s Cancer Group protocols from the 1990s. Control DNA samples from the Children’s Cancer Group were derived from buccal swabs (obtained for the creation of an anonymized genetic resource) from 162 unrelated persons of European descent living in the Los Angeles area during the period from 2000 through 2001.

Written informed consent was obtained from all participants, and the study was approved by each participating center’s institutional review board as well as by the Scientific Council and the Neuroblastoma Disease Committee of the Children’s Oncology Group and the Cancer Therapy Evaluation Program at the National Cancer Institute.

**GENOTYPING**

Details of methods for genomewide genotyping have been described elsewhere. Descriptions of these methods, along with methods for replication genotyping by means of polymerase chain reaction–based allelic discrimination assays, are included in the Supplementary Appendix, available with the full text of this article at www.nejm.org.

**STATISTICAL ANALYSIS**

Genomewide genotyping data from an initial 1251 patients with neuroblastoma and 2236 disease-free control subjects in the discovery phase were filtered on the basis of prespecified quality-control measures. Individual SNPs were excluded from further analysis if they showed deviation from the Hardy–Weinberg equilibrium with a P value of less than 0.001, an individual SNP genotype yield of less than 98%, or a minor allele frequency of less than 5%. This filtering resulted in the use of 464,934 SNPs in the subsequent analyses. A total of 33 samples (from 23 case subjects and 10 control subjects) had genotype yields of less than 90% and were removed. Because the case samples were accrued nationwide, whereas the control set was recruited locally in Philadelphia, we performed principal-components analyses to identify outlier samples to reduce the effects of population stratification. These analyses resulted in the removal of 379 samples (from 196 case subjects and 183 control subjects), resulting in 1032 patients and 2043 control subjects for our discovery-phase case series. Evaluation of these 3075 subjects with the use of ancestry informative markers available on the HumanHap550 BeadChip indicated European ancestry in all but 2 subjects; these 2 subjects remained in the analysis. The patients with neuroblastoma were representative of the expected distribution of clinical and biologic covariates as observed in patients with neuroblastoma in the general population (Table S1 in the Supplementary Appendix).

The primary statistical tests for association in the discovery-phase case series were carried out with the use of the PLINK software package. We conservatively set $1.0 \times 10^{-7}$ as the threshold for genomewide significance, on the basis of the fact that slightly less than 500,000 SNPs were used in the analysis ($0.05 \div 500,000 = 1.0 \times 10^{-7}$). The single-marker analyses for the genomewide data were carried out with the use of the chi-square test on the basis of differences in allele counts between 1032 case subjects and 2043 control subjects and the Cochran–Armitage test for trends on genotype frequencies. Allelic odds ratios and the corresponding 95% confidence intervals were calculated for the association analyses. In addition, to further control for the potential confounding influence of population stratification, we performed association analyses after correction for substructure based on a principal-components analysis as implemented in Eigenstrat. For each SNP, we used the default settings within the program and performed a modified Cochran–Armitage trend test adjusting for the top five principal components, and we report the result as the Eigenstrat P value. Figure S1 in the Supplementary Appendix shows quantile–quantile plots before and after correction.
Figure 1. Results of the Discovery Phase of the Neuroblastoma Genomewide Association Study.
The y axis represents the level of significance for each single-nucleotide polymorphism (log-transformed P values) at the relative genomic position on each chromosome along the x axis from the short-arm terminus (left) to the long-arm terminus (right). The black line indicates the genomewide significance threshold.

Table 1. Results of the Discovery Phase of a Neuroblastoma Genomewide Association Study for Case and Control Subjects.*

<table>
<thead>
<tr>
<th>Chromosome</th>
<th>SNP</th>
<th>Genome Position</th>
<th>Gene Reference Allele (A)</th>
<th>Minor Allele (B)</th>
<th>No. Case Subjects</th>
<th>Minor Allele Frequency</th>
<th>No. Control Subjects</th>
<th>Minor Allele Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>rs6939340</td>
<td>22247733</td>
<td>FLJ22536 A</td>
<td>G</td>
<td>1012</td>
<td>0.5553</td>
<td>2033</td>
<td>0.4715</td>
</tr>
<tr>
<td>6</td>
<td>rs4712653</td>
<td>22233943</td>
<td>FLJ22536 T</td>
<td>C</td>
<td>996</td>
<td>0.5387</td>
<td>2033</td>
<td>0.4555</td>
</tr>
<tr>
<td>6</td>
<td>rs9295536</td>
<td>22239908</td>
<td>FLJ22536 C</td>
<td>A</td>
<td>1031</td>
<td>0.5082</td>
<td>2040</td>
<td>0.4272</td>
</tr>
<tr>
<td>20</td>
<td>rs3790171</td>
<td>19170336</td>
<td>SLC24A3 T</td>
<td>C</td>
<td>1030</td>
<td>0.1898</td>
<td>2043</td>
<td>0.1361</td>
</tr>
<tr>
<td>20</td>
<td>rs7272481</td>
<td>19206143</td>
<td>SLC24A3 A</td>
<td>G</td>
<td>1030</td>
<td>0.1694</td>
<td>2043</td>
<td>0.1189</td>
</tr>
</tbody>
</table>

* Results shown are for single-nucleotide polymorphisms (SNPs) yielding a genomewide significant association, sorted according to unadjusted P value. Statistical tests for association were carried out with the use of the PLINK software package (http://pngu.mgh.harvard.edu/purcell/plink/). The single-marker analysis for the genomewide data was carried out with the use of a chi-square test on allele count differences between case subjects and control subjects. Results of the Cochran–Armitage test for trends and Eigenstrat analysis are also included. Allelic odds ratios and the corresponding 95% confidence intervals were calculated for the association analysis and are shown. Odds ratios are shown for the allelic test, as well as for heterozygosity (AB–AA) and homozygosity (BB–AA) for the B risk allele. Genome positions are based on the National Center for Biotechnology Information database, build 36 (hg18).
Of the 1032 patients included in the discovery-phase case series, clinical and biologic covariate data obtained at diagnosis were available for most (Table S1 in the Supplementary Appendix). Complete outcome data were available for 883 patients (85.6%), with a median follow-up interval of 4.02 years for patients without an event. Association analyses of chromosome 6p22 SNPs with clinical characteristics were performed with the chi-square test on allele and genotype counts. These analyses were also performed for outcome by comparing Kaplan–Meier survival curves by means of the log-rank test in a pairwise fashion.

**RESULTS**

To identify sequence variants that are associated with susceptibility to the development of neuroblastoma, we compared single-marker allele and genotype frequencies in our discovery-phase case series, using statistics based on chi-square analysis and the Cochran–Armitage trend test. The Eigenstrat-corrected summary statistics for the full data set are available in the repository of the National Institutes of Health Genotype and Phenotype database (dbGAP; www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000124.v1.p1) (accession number, phs000124.v1.p1). The top three SNPs showing a significant association to neuroblastoma were in close linkage disequilibrium at chromosome 6p22 (rs6939340, rs4712653, and rs9295536) yielding P values of $7.01 \times 10^{-10}$ to $1.71 \times 10^{-9}$ (allelic odds ratio, 1.39 to 1.40) (Fig. 1 and Table 1, and Table S2 in the Supplementary Appendix). Two additional SNPs at chromosome 20p11 (rs790171 and rs7272481) showed genomewide significant single-marker P values, and many others were very close to the genomewide significance threshold (Fig. 1). However, only the chromosome 6 association-study results retained genomewide significance after further correction for population substructure with the use of principal-components analyses as implemented in Eigenstrat\textsuperscript{22,23} (Table 1). The signal on chromosome 6 falls within a 94.2-kb linkage-disequilibrium block containing the predicted overlapping genes FLJ22536 and FLJ44180 (Fig. 2).

We next sought to replicate the association signals of chromosomes 6p22 and 20p11 in three separate pairings of patients with neuroblastoma and control subjects. As shown in Table 2, the risk alleles in chromosome 6 identified in the discovery phase were also significantly overrepresented in all three groups of patients with neuroblastoma as compared with control subjects, yielding a combined P value and allelic odds ratio for the most strongly associated SNP rs6939340 of $3.64 \times 10^{-5}$ and 1.49, respectively (95% confidence interval, 1.29 to 1.71). We did not observe replication of the association between the alleles in chromosome 20 and neuroblastoma (identified in the discovery cohort) in the three replication case series (Table S3 in the Supplementary Appendix).
To determine whether the three implicated SNPs at the chromosome 6p22 locus were differentially associated with patient subgroups and outcome, we analyzed their prevalence according to prognostically relevant clinical and biologic covariates present at diagnosis and survival rates among patients in the case series from the Children’s Oncology Group. All three SNPs were associated with a more malignant clinical presentation and aggressive disease course (Fig. 3). Accordingly, tumors that were metastatic at diagnosis were more likely to develop in patients with neuroblastoma who were homozygous for the at-risk alleles (P=0.006, P=0.008, and P=0.024 for rs4712653, rs9295536, and rs6939340, respectively), to have somatically acquired amplification.
of the MYCN oncogene in tumor cells (P = 0.002, P = 0.003, and P = 0.006, respectively), and to have a high-risk classification for the purposes of treatment stratification (P = 0.002, P = 0.001, and P = 0.011, respectively), as compared with patients with neuroblastoma who were homozygous for the alleles that were not associated with risk (Table S4 and S5 in the Supplementary Appendix). In addition, patients who were homozygous for the risk alleles had a significantly decreased probability of event-free survival (P = 0.016 for SNP rs6939340) (Fig. S2 in the Supplementary Appendix). This association with more malignant disease may explain why we were able to show replication of results in the relatively small series of 59 patients from the Children's Cancer Group that consisted solely of high-risk patients (Table 2).

### DISCUSSION

This study shows that the likelihood for malignant transformation of developing neuroblasts is influenced by variants on chromosome 6p22. Associations between neuroblastoma and other loci may come to light as we continue toward our goal of the genomewide genotyping of DNA samples from 5000 patients with neuroblastoma. Our data provide support, on a preliminary level, for the “common variant, common disease” model for neuroblastoma (i.e., the interaction of multiple relatively common genetic variations in the developing neuroblast may predispose toward the development of the disease). The motivation for this large and ongoing study was the paucity of information on neuroblastoma tumorigenesis. Our study provides proof-
of-concept results for a genomewide-association approach to neuroblastoma and identifies new candidate susceptibility genes, although little is known about these genes.

**FLJ22536** has multiple predicted isoforms and contains a potential epidermal growth factor–like domain. The **FLJ44180** has no sequence similarities in human or mouse nucleic acids and protein databases. Studies are under way to determine how the presence of common DNA variants at 6p22 contributes to the risk of neuroblastic malignant transformation and the subphenotype of high-risk disease.

Supported by the efforts of multiple investigators throughout the Children’s Oncology Group (U10-CA98543, Dr. Gregory Reaman, chair); grants (R01-CA78454, to Dr. Maris; R01-ES009911, to Dr. Li; and R01-CA60104, to Dr. Seeger) from the National Institutes of Health, the Alex’s Lemonade Stand Foundation (to Dr. Maris), the Center for Applied Genomics at the Joseph Stokes Jr. Research Institute, which funded all genomewide association genotyping (to Dr. Hakonarson), the Abramson Family Cancer Research Institute (to Dr. Maris), the Neuroblastoma Society (to Drs. Rahman and Scott), and the Institute of Cancer Research (to Drs. Rahman and Scott); a fellowship from the Associazione Italiana per la Ricerca sul Cancro (to Dr. Capasso); and the Factors Associated with Childhood Tumours Study and the Tumour and Leukaemia Bank of the Childhood Cancer and Leukaemia Group. DNA was provided by the British 1958 Birth Cohort collection, funded by grants from the Medical Research Council (G0000934) and from the Wellcome Trust (068545/Z/02).

No potential conflict of interest relevant to this article was reported.
REFERENCES