Common variation at 6q16 within \textit{HACE1} and \textit{LIN28B} influences susceptibility to neuroblastoma

Sharon J Diskin$^{1–3,11}$, Mario Capasso$^{4,5}$, Robert W Schnep$^{1,2}$, Kristina A Cole$^{1–3}$, Edward F Attiyeh$^{1–3}$, Cuiping Hou$^6$, Maura Diamond$^{1,2}$, Erica L Carpenter$^{1,2}$, Cynthia Winter$^{1,2}$, Hanna Lee$^{1,2}$, Jayanti Jagannathan$^{1,2}$, Valeria Latorre$^{7,8}$, Achille Iolascon$^{4,5}$, Hakon Hakonarson$^{3,6,7}$, Marcella Devoto$^{3,7,9,10}$ & John M Maris$^{1–3,11}$

Neuroblastoma is a cancer of the sympathetic nervous system that accounts for approximately 10% of all pediatric oncology deaths$^1$. Here, we report a genome-wide association study of 2,817 neuroblastoma cases and 7,473 controls.

We identified two new associations at 6q16, the first within \textit{HACE1} (rs4336470; combined $P = 2.7 \times 10^{-11}$; odds ratio 1.26, 95% confidence interval (CI) 1.18–1.35) and the second within \textit{LIN28B} (rs17065417; combined $P = 1.2 \times 10^{-8}$; odds ratio 1.38, 95% CI 1.23–1.54). Expression of \textit{LIN28B} and \textit{let-7} miRNA correlated with rs17065417 genotype in neuroblastoma cell lines, and we observed significant growth inhibition upon depletion of \textit{LIN28B}, specifically in neuroblastoma cells that were homozygous for the risk allele. Low \textit{HACE1} and high \textit{LIN28B} expression in diagnostic primary neuroblastomas were associated with worse overall survival ($P = 0.008$ and 0.014, respectively). Taken together, these data show that common variants in \textit{HACE1} and \textit{LIN28B} influence neuroblastoma susceptibility and indicate that both genes likely have a role in disease progression.

Neuroblastoma is a malignancy derived from the developing sympathetic nervous system. The median age at diagnosis is 17 months, and the survival rate for the most aggressive subset remains approximately 50%, despite intensive multimodal cytotoxic therapy$^1$ and recent advances in immunotherapy$^2$. The genetic etiology of familial neuroblastoma, which accounts for approximately 1% of cases, has recently come into focus$^3–6$; however, the genetic and environmental factors that cause sporadic neuroblastoma remain largely unknown. We have recently reported common SNPs within or upstream of \textit{LINC00340} and \textit{LOC729177} (also known as \textit{FLJ44180}), \textit{BARD1}, \textit{LMO1}, \textit{DUSP12}, \textit{HSD17B12} and \textit{DDX4-IL31RA} and a common copy-number variation (CNV) within \textit{NBPF23} are each highly associated with neuroblastoma$^7–11$. Collectively, however, these variants still account for only a small portion of neuroblastoma heritability, and it is likely that additional predisposition loci remain to be discovered.

To identify additional variants associated with neuroblastoma, we expanded our previous genome-wide association study (GWAS) discovery cohort and analyzed 2,101 neuroblastoma cases accrued through the North America–based Children’s Oncology Group (Supplementary Table 1) with 4,202 control subjects of European ancestry who were matched genetically and by genotyping array version to minimize genomic inflation (Online Methods and Supplementary Figs. 1 and 2). All subjects were genotyped using the Illumina HumanHap550 or Quad610 BeadChips. We restricted our analysis to the SNPs present on both platforms that passed our quality control metrics; the genomic control inflation factor was 1.14 (Supplementary Fig. 3). Evaluation of the first three principal components in cases and controls confirmed that the slightly high inflation factor was not due to gross population stratification (Supplementary Fig. 4). Clusters of SNPs from six genomic loci reached genome-wide significance ($P$ values ranged from $7.8 \times 10^{-16}$ to $4.8 \times 10^{-8}$; Supplementary Fig. 1 and Table 2), including three SNPs within \textit{LINC00340} and \textit{LOC729177} (FLJ44180) at 6p22 ($P$ values ranged from $7.8 \times 10^{-16}$ to $1.7 \times 10^{-14}$), ten SNPs within or near \textit{BARD1} at 2q35 ($P$ values ranged from $4.1 \times 10^{-14}$ to $3.7 \times 10^{-8}$), two SNPs within \textit{LMO1} at 11p15 ($P$ values ranged from $1.2 \times 10^{-12}$ to $3.8 \times 10^{-10}$) and one SNP within \textit{HSD17B12} at 11p11 ($P = 4.8 \times 10^{-8}$), further confirming our previous reports$^7–10$. In addition, we identified one SNP (rs4696715) at chromosome 4p16 and two SNPs (rs4336470 and rs9404576) at 6q11 that have not been reported previously ($P$ values ranged from $1.8 \times 10^{-8}$ to $3.4 \times 10^{-8}$; Table 1). Several genotyped SNPs in strong linkage disequilibrium (LD) with rs4696715 at 4p16 did not show evidence for association with neuroblastoma; therefore, this SNP was not considered further.

Closer examination of the 6q16 locus identified four additional SNPs that showed association ($P < 1 \times 10^{-4}$) with neuroblastoma.

---

1Division of Oncology, Children’s Hospital of Philadelphia, Philadelphia, Pennsylvania, USA. 2Center for Childhood Cancer Research, Children’s Hospital of Philadelphia, Philadelphia, Pennsylvania, USA. 3Department of Pediatrics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA. 4Dipartimento di Biochimica Biotecnologie Mediche, Università degli Studi di Napoli Federico II, Naples, Italy. 5Ceinge–Biotecnologie Avanzate, Naples, Italy. 6The Center for Applied Genomics, Children’s Hospital of Philadelphia, Philadelphia, Pennsylvania, USA. 7Division of Genetics, Children’s Hospital of Philadelphia, Philadelphia, Pennsylvania, USA. 8Department of Cell Biology, Università della Calabria, Rende, Italy. 9Department of Molecular Medicine, University of Rome La Sapienza, Rome, Italy. 10Department of Biostatistics and Epidemiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA. 11Abramson Family Cancer Research Institute, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA. Correspondence should be addressed to J.M.M. (maris@chop.edu).

Received 16 March; accepted 24 July; published online 2 September 2012; doi:10.1038/ng.2387
Table 1 Significantly associated genotyped SNPs at the HACE1 and LIN28B loci at 6q16

| SNP          | Discovery cohorta | Italian replicationa | African-American replicationa | Combined                |
|--------------|-------------------|----------------------|-----------------------------|-------------------------|
|              | n (A1/A2) | Freq. A1 | Freq. A1 | n (A1/A2) | Freq. A1 | Freq. A1 | n (A1/A2) | Freq. A1 | Freq. A1 | n (A1/A2) | Freq. A1 | Freq. A1 | Pvalue | OR (95% CI) |
| rs4336470    | 2,101 | 0.30    | 0.35    | 351  | 0.30    | 0.34    | 370  | 0.257 | 0.40    | 1.8 × 10⁻⁸ | 0.060 | 0.70 | 1.4 × 10⁻³ | 2.7 × 10⁻⁷ | 1.26 (1.18–1.35) |
| rs9404576    | 4,202 | 0.30    | 0.35    | 780  | 0.30    | 0.34    | 790  | 0.257 | 0.40    | 3.4 × 10⁻⁸ | 0.060 | 0.70 | 1.3 × 10⁻³ | 1.8 × 10⁻⁷ | 1.27 (1.18–1.36) |
| rs4079063    | 365   | 0.43    | 0.47    | 365  | 0.43    | 0.47    | 365  | 0.257 | 0.40    | 4.0 × 10⁻⁸ | 0.060 | 0.70 | 1.3 × 10⁻³ | 1.3 × 10⁻⁷ | 1.20 (1.12–1.29) |
| rs2499663    | 1,000 | 0.43    | 0.47    | 463  | 0.43    | 0.47    | 463  | 0.257 | 0.40    | 4.5 × 10⁻⁸ | 0.060 | 0.70 | 1.5 × 10⁻⁴ | 1.6 × 10⁻⁷ | 1.21 (1.13–1.29) |
| rs2499667    | 365   | 0.43    | 0.47    | 365  | 0.43    | 0.47    | 365  | 0.257 | 0.40    | 2.6 × 10⁻⁸ | 0.060 | 0.70 | 2.6 × 10⁻⁴ | 1.2 × 10⁻⁷ | 1.21 (1.13–1.29) |

rs17065417 C/A 0.08 0.11 1.8 × 10⁻³ 0.08 0.11 0.033 0.09 0.11 0.129 1.2 × 10⁻⁸ 1.38 (1.23–1.54)

(County Children’s Hospital of Philadelphia; freq., frequency; *P values were calculated by genotypic test. **P values were calculated by logistic regression with percent African admixture as covariate.)

Meta-analysis P value calculated using METAL. OR, odds ratio of risk allele based on meta-analysis.

(1 and Supplementary Table 3) These of the HACE1 gene (encoding HECT domain– and ankyrin repeat–containing E3 ubiquitin protein ligase 1) and exhibited a moderate degree of LD with rs4336470. The fourth SNP (rs17065417; \( P = 1.8 \times 10⁻⁸ \)) mapped to an intron of the LIN28B gene (encoding-lin28 homolog B) and showed very little evidence for LD with rs4336470 in multiple HapMap populations (Supplementary Fig. 5). To ensure that these results were not influenced by subtle substructure in the discovery phase, we included the first 20 principal components as covariates in a logistic regression analysis. This reduced the inflation factor to 1.04 but did not alter our conclusions regarding regions that were associated at genome-wide significance, and the P values for SNPs at the newly identified 6q16 locus were essentially unchanged (Supplementary Table 4). Finally, we conditionally the analysis of the 6q16 SNPs on rs4336470 to investigate whether there might be more than one independent association signal at 6q16. As expected, the associations for SNPs in modest LD with rs4336470 were no longer statistically significant after conditioning, and these SNPs clearly represent one signal (Supplementary Table 5). In contrast, whereas the signal at rs17065417 was attenuated, it remained significant (\( P = 2.5 \times 10⁻⁴ \); Supplementary Table 5), suggesting that rs4336470 and rs17065417 may contribute independently to neuroblastoma risk. No association was observed between rs4336470 or rs17065417 genotype and clinical and biological covariates (Supplementary Tables 6 and 7), and we observed only weak evidence for epistasis at the other significant loci (Supplementary Tables 8 and 9), suggesting that each may act independently to confer risk.

We next sought to replicate the rs4336470 and rs17065417 associations in an Italian cohort of 351 cases and 780 controls using PCR-based genotyping. Both SNPs showed evidence for association in the same direction as seen in the discovery effort (Table 1 and Supplementary Table 3). To assess whether these variants influence susceptibility in other ancestry groups and to seek additional replication, we analyzed a third independent case series comprised of 365 African-American neuroblastoma cases and 2,491 genet Rochester Institute of Medicine; freq., frequency. \( P \) values were calculated by logistic regression with percent African admixture as covariate.)

Meta-analysis P value calculated using METAL. OR, odds ratio of risk allele based on meta-analysis.

(Online Methods and Supplementary Fig. 6). Consistent with the lower incidence rate of neuroblastoma in African-Americans, the rs4336470[T] protective allele is actually the major allele in African-Americans, whereas it is the minor allele in individuals of European ancestry. Despite these observed differences in allele frequencies, all SNPs mapped to HACE1 replicated robustly in the African-American cohort (\( P \) values ranged from \( 1.3 \times 10⁻⁷ \) to \( 1.4 \times 10⁻³ \); Table 1 and Supplementary Table 3), and all associations were in the same direction. Allele frequencies for the LIN28B SNP rs17065417 were comparable across populations of different ancestry and showed a trend toward association in African-Americans in the same direction as in the European-American discovery and Italian replication cohorts (Table 1 and Supplementary Table 3). Combined analysis using the weighted inverse-variance method in METAL showed that all six genotyped SNPs had \( P \) values beyond or approaching the conservative Bonferroni-adjusted genome-wide
Figure 2 LIN28B risk alleles correlate with increased LIN28B expression and decreased let-7 miRNA expression. (a) LIN28B mRNA expression is significantly higher in neuroblastoma cell lines homozygous for the rs17065417(A) risk allele (AA) compared to neuroblastoma cell lines heterozygous for the risk allele (CA). Given the MAF of rs17065417, we did not identify any cell lines homozygous for the rs17065417(C) protective allele. (b) Neuroblastoma cell lines homozygous for the rs17065417 risk allele show high expression of MYCN. (c) Protein blot and corresponding densitometry analysis confirms higher amounts of LIN28B protein in neuroblastoma cell lines homozygous for the rs17065417 risk allele. (d) Neuroblastoma cell lines homozygous for the rs17065417 risk allele that have high LIN28B expression show decreased or absent expression of the let-7 miRNAs.

To identify variants at the 6q16 locus that influence neuroblastoma cell growth in an expression-specific manner, we performed genotype imputation in our discovery cohort using data from the 1000 Genomes Project (Fig. 1). Five of these imputed SNPs were in strong LD with rs4336470 (r^2 > 0.8 in the 1000 Genomes Project European (EUR) population); three mapped to introns of HACE1 and the other two were located just downstream of HACE1 (Fig. 1a). The remaining five imputed SNPs reaching genome-wide significance were in strong LD with rs17065417 (r^2 > 0.8 in the 1000 Genomes Project EUR population) and were located within introns of LIN28B (Fig. 1b). In total, we identified 46 imputed SNPs that showed strong evidence for association (P < 1.0 × 10^{-6}; Supplementary Table 10). To further evaluate whether two independent association signals exist at 6q16, we conditioned the regional association analysis on either rs4336470 or rs17065417, finding that neither could fully account for the observed associations (Supplementary Fig. 7). The signal at 6q16 was only abolished across the entire region after conditioning on both rs4336470 and rs17065417 (Supplementary Fig. 8). Although we cannot rule out the possibility that both SNPs may be tagging the same underlying risk variant, these data are consistent with the presence of two independent association signals at 6q16, one implicating HACE1 and the other LIN28B.

The HACE1 gene encodes an E3 ubiquitin protein ligase that was first identified in a sporadic Wilms’ tumor that harbored a t(6;15)(q21;q21) translocation. The rearrangement was associated with decreased HACE1 expression, and further study showed that...
**Figure 4** *HACE1* and *LIN28B* expression are associated with advanced disease and poor outcome in neuroblastoma. (a,b) mRNA expression for Children’s Oncology Group (COG) risk groups of *HACE1* (a) and *LIN28B* (b). The number of tumors in each group is indicated in parentheses. Error bars, s.e.m. NBL, neuroblastoma. (c,d) Kaplan-Meier analysis is shown, with individuals grouped by tertiles of expression of *HACE1* (c) and *LIN28B* (d). Log-rank *P* values are shown. (e,f) Replication of changes in expression for *HACE1* (e) and *LIN28B* (f) in advanced-stage neuroblastoma using published Affymetrix U133 plus v2 array data (R2 bioinformatics tool). Data are shown for International Neuroblastoma Staging System (INSS) stages 1–4. The number of tumors in each group is indicated in parentheses. Error bars, s.e.m. (g,h) Kaplan-Meier analysis is shown, with individuals grouped by tertiles of expression of *HACE1* (g) and *LIN28B* (h). Log-rank *P* values are shown. Data sets for g and h correspond to those in e and f, respectively. **** *P* < 0.0001; *** *P* < 0.001; ** *P* < 0.01; * *P* < 0.05.

*HACE1* is silenced in the majority of Wilms’ tumors via hypermethylation of two CpG islands upstream of the transcriptional start site\(^\text{15}\). Similar epigenetic silencing has been reported in advanced colorectal cancer\(^\text{16}\) and gastric carcinoma\(^\text{17}\). Indeed, *HACE1* is downregulated in multiple human tumors and maps to a region of common deletion or loss of heterozygosity (LOH), consistent with a tumor suppressor function. *Hace1*-null mice form spontaneous tumors in a wide array of tissues and are susceptible to additional cancer triggers, both genetic and environmental\(^\text{18}\). Through its E3 ubiquitin ligase function, *HACE1* has been shown to suppress cell growth and anchorage independence of human tumor cells, including in the neuroblastoma cell line IMR32 (ref. 18). Studies suggest that *HACE1* inhibits cell cycle progression during stress via regulation of cyclin D1 degradation\(^\text{18}\) and that *HACE1* also regulates retinoic acid receptor (RAR) activity\(^\text{19}\). In addition to multiple somatic alterations involving the *HACE1* locus, a constitutional t(5;6)(q21;q21) translocation that disrupts *HACE1* was recently identified, making it a putative Wilms’ tumor susceptibility gene\(^\text{20}\).

The *LIN28B* gene encodes a developmentally regulated RNA-binding protein and is a key repressor of the let-7 family of microRNAs (miRNAs)\(^\text{21}\). Both *LIN28A* and *LIN28B* are oncogenes that promote cellular transformation when ectopically overexpressed\(^\text{22–24}\), and high levels of expression of *LIN28A* or *LIN28B* have been observed in several human cancers and correlate with low let-7 levels\(^\text{22,24}\). Post-transcriptional regulation of let-7 by *LIN28A* is required for normal development and contributes to the pluripotent state by preventing let-7–mediated differentiation of embryonic stem cells\(^\text{21,25,26}\). Overexpressing *LIN28A* or *LIN28B* or inhibiting let-7 with antisense RNA promotes reprogramming of human and mouse fibroblasts into pluripotent stem cells\(^\text{26,27}\). In a panel of 60 pediatric cancer cell lines\(^\text{28}\), *LIN28B* was consistently expressed at high levels in neuroblastoma (Supplementary Fig. 9). Recently, *LIN28B* and let-7 were also identified as key regulators of glucose homeostasis\(^\text{29}\) and hematopoiesis\(^\text{30}\).

To date, GWAS have identified variants in *LIN28B* associated with human height\(^\text{31}\) and the ages of onset of puberty\(^\text{32}\) and menarche\(^\text{33}\). In addition, a recent candidate gene study identified a putative association with epithelial ovarian cancer\(^\text{34}\).

To investigate the functional relevance of neuroblastoma-associated SNPs within *HACE1* and *LIN28B*, we first analyzed a set of 12 neuroblastoma cell lines with matched genome-wide SNP genotyping and mRNA expression data. No correlation between rs4336470 and *HACE1* mRNA expression was observed (*P* = 0.30, but *LIN28B* expression was significantly higher in cell lines homozygous for the rs17065417[A] risk allele compared to heterozygous cell lines (*P* = 0.02; Fig. 2a). No cell lines tested were homozygous for the rs17065417[C] protective allele. *LIN28A* was not expressed in neuroblastoma cell lines, consistent with recent reports that *LIN28A* and *LIN28B* may be mutually exclusive in terms of their expression in cancer cells\(^\text{21}\). In addition to a correlation between rs17065417 genotype and *LIN28B* expression, we observed a strong positive correlation with MYCN expression (*P* = 0.0003; Fig. 2b). We next confirmed *LIN28B* protein levels by protein blot in four cell lines at the extremes of *LIN28B* mRNA expression (Fig. 2c). Cell lines expressing high amounts of *LIN28B* showed lower let-7 expression across the entire miRNA family (Fig. 2d). Transient knockdown of *LIN28B* resulted in significant growth inhibition in neuroblastoma cells homozygous for the rs17065417 risk allele that had high *LIN28B* expression (Fig. 3a–d) but not in the heterozygous cell line SKNAS that had low *LIN28B* expression (Fig. 3e,f). Together, these data are consistent with the hypothesis that *LIN28B* promotes neuroblastoma tumorigenesis, in part through repression of let-7 family miRNAs, and that the risk alleles are associated with growth advantage through increased *LIN28B* expression; however, additional studies with larger sample sizes are needed to confirm this model.

To examine the relevance of *HACE1* and *LIN28B* in tumor samples, we assayed expression of both genes by quantitative RT-PCR in a representative set of 87 primary tumors obtained at time of diagnosis (Supplementary Table 1). *HACE1* expression was significantly lower (*P* = 0.002; Fig. 4a), and *LIN28B* expression was significantly higher (*P* = 0.032; Fig. 4b) in the high-risk tumors compared to the...
low-risk tumors. Accordingly, low HACE1 expression was associated with worse overall survival (P = 0.008; Fig. 4c), as was high LIN28B expression (P = 0.015; Fig. 4d). Analysis of four independent mRNA expression array data sets comprising 517 neuroblastoma tumors provided robust replication of these observations (Fig. 4e–h and Supplementary Figs. 10–12). These data support the hypothesis that HACE1 may function as a tumor suppressor and LIN28B as an oncogene in advanced neuroblastomas.

In conclusion, we have identified common variants within HACE1 and LIN28B that are associated with neuroblastoma. As with other genes identified in our ongoing GWAS efforts,8,9,35 it is likely that the germline susceptibility variants identified here have an important role, not only in tumor initiation, but also in disease progression via cis effects on major cancer-associated genes. Further study of HACE1 and LIN28B in neuroblastoma is warranted and may lead to new insights into the genetic and epigenetic mechanisms underlying an aggressive clinical phenotype.

URLs. Database of Genotypes and Phenotypes (dbGaP), http://www.ncbi.nlm.nih.gov/gap; LocusZoom, http://csg.sph.umich.edu/locuszoom/; 1000 Genomes Project, http://www.1000genomes.org/; R2 bioinformatics tool, http://hgserver1.amc.nl/cgi-bin/r2/main.cgi; National Cancer Institute (NCI) Oncogenomics, http://home.ccr.cancer.gov/oncology/oncogenomics/; LiftOver tool, http://genome.ucsc.edu/cgi-bin/hgLiftOver; 1000 Genome Project Phase 1 interim data, http://mathgen.stats.ox.ac.uk/impute/data_download_1000G_phase1_interim.html.

METHODS

Methods and any associated references are available in the online version of the paper.

Accession codes. Genotyping data have been deposited at dbGaP under accession phs000124.

Note: Supplementary information is available in the online version of the paper.

ACKNOWLEDGMENTS

We thank P. Sleiman for useful discussions regarding meta-analysis using METAL. This work was supported in part by US National Institutes of Health (NIH) grants R01-CA124709 (J.M.M.), K99-CA151869 (S.J.D.), P30-HD026979 (M. Devoto) and K08-CA136979 (K.A.C.); the Giulio D’Angio Endowed Chair (J.M.M.); the Alex’s Lemonade Stand Foundation (J.M.M.); Andrew’s Army Foundation (J.M.M.); the PressOn Foundation (J.M.M.); the Abramson Family Cancer Research Institute (J.M.M.); Fondazione Italiana per la Lotta al Neuroblastoma and Associazione Italiana per la Ricerca sul Cancro (M.C.); and the Center for Applied Genomics at the Children’s Hospital of Philadelphia Research Institute (H.H.).

AUTHOR CONTRIBUTIONS

S.J.D. and J.M.M. designed the experiment and drafted the manuscript. S.J.D. analyzed SNP data, performed SNP association studies, and analyzed mRNA and miRNA expression data. M.C. and A.I. replicated SNP associations in the African-American cohort. E.I.C. and H.H. confirmed LIN28B protein expression by protein blot. R.W.S and C.W. performed experiments with siRNA-mediated knockdown of LIN28B. E.F.A. generated miRNA expression array data, including low-level summary values. K.A.C. performed RT-PCR in primary tumors. M. Diamond and C.H. organized samples and genotyped cases. J.I. participated in expression array analyses. H.H. generated and provided all control data for the GWAS. M. Devoto and H.H. contributed to overall study design. All authors commented on or contributed to the current manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at http://www.nature.com/doifinder/10.1038/ng.2387. Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Maris, J.M. Recent advances in neuroblastoma. N. Engl. J. Med. 362, 2202–2211 (2010).
2. Yu, A.L. et al. Anti-GD2 antibody with GM-CSF, interleukin-2, and isotretinoin for neuroblastoma. N. Engl. J. Med. 363, 1324–1334 (2010).
3. Mossé, Y.P. et al. Identification of ALK as a major familial neuroblastoma predisposition gene. Nature 455, 930–935 (2008).
4. Jornuex-Lerosey, I. et al. Common variants in BARD1 influence susceptibility to high-risk neuroblastoma. Nature 455, 967–970 (2008).
5. Trochet, D. et al. Germline mutations of the paired-like homeobox 2B (PHOX2B) gene in neuroblastoma. Am. J. Hum. Genet. 74, 761–764 (2004).
6. Mosse, Y.P. et al. Germline PHOX2B mutation in hereditary neuroblastoma. Am. J. Hum. Genet. 75, 727–730 (2004).
7. Maris, J.M. et al. Chromosome 6p22 locus associated with clinically aggressive neuroblastoma. N. Engl. J. Med. 358, 2585–2593 (2008).
8. Capasso, M. et al. Common variants in BARD1 associated with age of diagnosis. Hum. Mol. Genet. 21, 31–35 (2012).
Cases were defined as subjects who were diagnosed with neuroblastoma or ganglioneuroblastoma and registered through COG. All specimens were obtained at time of diagnosis, and the majority were annotated with clinical and genomic information that included age at diagnosis, site of origin, INSS disease stage, International Neuroblastoma Pathology Classification (INPC), MYCN oncogene copy number, DNA index (ploidy), registration in clinical trials, and overall survival and second malignancies.

Control subjects were recruited from the Philadelphia region through the CHOP 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 (i) self-reporting as Caucasian; (ii) availability of 1.5 μg of high-quality DNA from peripheral blood mononuclear cells; and (iii) no serious underlying medical disorder, including cancer. The Research Ethics Board of the Children's Hospital of Philadelphia (CHOP) approved the study, and written informed consent was obtained from all subjects by nursing and medical assistant staff under the direction of CHOP clinicians.

Genome-wide SNP genotyping. Genotyping for both discovery and replication phases was performed using the Illumina Infinium II HumanHap550 and Human Quad610 BeadChips according to methods detailed elsewhere (34, 42-43) and summarized here. DNA samples were surveyed for quality, both by optical density spectrophotometry and PicoGreen assay (Invitrogen). A total of 750 ng of DNA isolated from blood was used to genotype each sample according to the manufacturer’s guidelines. On day 1, genomic DNA was amplified 1,000–1,500-fold. On day 2, amplified DNA was fragmented to ~300–600 bp in size and was precipitated and resuspended. Fragments were then hybridized to the BeadChip arrays. Single-base extension (SBE) used a single probe sequence of approximately 50 bp designed to hybridize immediately adjacent to the SNP query site. Following targeted hybridization to the array, locus-specific primers for the arrayed SNP (attached to the beads) were extended with a single haptened dyeoxynucleotide in the SBE reaction. The haptens were subsequently detected by a multilayer immunohistochemical sandwich assay. The Illumina BeadArray Reader scanned each BeadChip at two wavelengths and created an image file. As BeadChip images were collected, intensity values were determined for all instances of each bead type, and data files were created that summarized intensity values for each bead type. These files consisted of intensity data that were loaded directly into the Illumina GenomeStudio genotype analysis software. Once normalization was complete, the clustering algorithm was run to evaluate cluster positions for each locus and assign individual genotypes. Only samples yielding a genotype call rate of ≥95% were considered for inclusion in this study.

Quality control and association testing for the discovery cohort. Overlap of the HumanHap550 v1, HumanHap550 v3 and Quad610 arrays. The discovery cohort consisted of individuals genotyped on the HumanHap550 v1, HumanHap550 v3 and the Human Quad610 arrays. Our analysis only considered markers shared by all three arrays. The HumanHap550 v1 array contains 555,175 markers, the HumanHap550 v3 array contains 561,288 markers and the Quad610 array contains 620,901 markers. Overall, 535,752 markers are shared by all three arrays and were analyzed in this study.

Low genotype call rate (<95%). Call rate was calculated on the basis of the number of ‘no call’ genotypes, with default genotyping calling algorithm implemented in Illumina GenomeStudio software. We did not consider any sample with a call rate of <95% for inclusion in this study. The call rate per individual was assessed by PLINK software and was confirmed to exceed 95% for all individuals, with an average genotyping rate of 99.85% across included individuals.

Inferring population structure for individuals of European ancestry. Multidimensional scaling (MDS), as implemented in PLINK software, was used to infer population structure in the neuroblastoma data set. Comparing self-identified ancestry with MDS-inferred ancestry confirmed the reliability of MDS in identifying genetically inferred individuals of European ancestry (Supplementary Fig. 2).

Detection and elimination of cryptic relatedness and duplicate genotypes. To detect cryptic relatedness and potential duplicate genotypes within our data, we applied a two-step procedure to calculate pairwise identity-by-descent (IBD) estimates between all individual case and control subjects. First, we examined the initial MDS and retained only those individuals of inferred European ancestry with call rates greater than 95%. Second, we recalculated genome-wide identity-by-state (IBS) estimates and IBD estimates in the remaining individuals of European ancestry using PLINK software. This two-step procedure ensures that allele frequency differences between populations do not lead to biases in IBD estimation. We applied a stringent threshold for detecting cryptic relatedness: any pair of subjects with IBD of >0.15 were flagged, and one individual was removed, so that only unrelated subject remained in the final association test. A total of 11 neuroblastoma cases were excluded.

Low call rate per marker (<99%). Markers with call rates of less than 99% were excluded from analysis. Call rates were calculated by PLINK software. A total of 49,130 markers were excluded from association analysis in this step.

Minor allele frequency (individuals of European ancestry). Markers with minor allele frequency (MAF) of less than 5% were excluded from our analysis. MAFs were calculated by PLINK software. A total of 50,159 markers were excluded from association analysis in this step.

Hardy-Weinberg equilibrium (individuals of European ancestry). Markers with Hardy-Weinberg equilibrium P values of less than 0.001 were excluded from analysis. A total of 11,953 markers were excluded from association analysis in this step.

Quality control and replication in the African-American cohort. The quality control procedure for the African-American replication cohort is largely similar to those performed on the discovery cohort and is described in detail elsewhere (37). The replication cohort after quality control consisted of a total of 365 cases and 2,491 control subjects who were genotyped on the Illumina HumanHap550 or Human610-Quad BeadChips. Significant SNPs from the discovery effort (P < 1 × 10−8) at the 6q16 loci were tested for association with neuroblastoma by logistic regression, including admixture estimate as a covariate.

Replication in Italian cohort. Genotyping of the top SNPs at HACE1 (rs4336470) and LIN28B (rs17065417) was performed using TaqMan SNP genotyping assays (Life Technology) for 350 neuroblastoma cases and 780 controls. Twenty DNA samples were also genotyped by Sanger sequencing for a further validation.

Meta-analysis. Meta-analysis was performed using the inverse-variance method within the METAL software package, and a fixed-effects model was assumed. The Cochran Q test was used to assess evidence of between-study heterogeneity of effect sizes.

Genotype imputation. Imputation was performed with IMPUTE2 (ref. 44) using the worldwide 1000 Genomes Project Phase 1 interim data as reference (June 2011 release). Rather than preselecting a reference population, we elected to follow the approach of Howie et al. and use a multilocus reference panel with IMPUTE2. Genotypes for markers located on chromosome 6 were extracted and mapped to the hg19 human reference genome using the LiftOver tool. Multipopulation haplotype data from the 1000 Genomes Project Phase 1 interim release (June 2011) were downloaded from the 1000 Genomes Project website (see URLs). IMPUTE2 was applied with default parameters and
Ne = 20,000. Following imputation, SNPs with MAF of <1% and/or IMPUTE2 info quality score of <0.8 were removed. To correct for uncertainty in the data resulting from the imputation process, the remaining SNPs were tested for association with neuroblastoma using the frequentist association test under the additive model implemented in SNPTTEST.

**Genome-wide mRNA expression profiling of neuroblastoma cell lines.**

Genome-wide mRNA expression profiling in neuroblastoma cell lines was performed using the Illumina WG-6 expression array according to the manufacturer’s specifications. Data were normalized using the average normalization method provided in Illumina GenomeStudio software. Two-sided t-tests were performed at the gene level to assess differential expression in cell lines. P < 0.05 was considered significant.

**miRNA expression profiling of neuroblastoma cell lines.**

RNA from samples was labeled with oligonucleotides from the Illumina MicroRNA Expression Profiling Assay (version 1) and then hybridized to a universal Sentrix Array Matrix according to the manufacturer’s specifications. Average signal values were normalized (rank invariant) using Illumina GenomeStudio software.

**LIN28B protein detection.**

Neuroblastoma cell lines were grown in T75 flasks under standard cell culture conditions. Whole-cell lysates were extracted with 100 µl of cell extraction buffer (Invitrogen) containing protease inhibitors (Sigma) and phenylmethyl sulfonl fluoride, were briefly sonicated and were rotated for 1 h at 4°C. After 30 min of centrifugation at 4°C, the supernatant was removed, and protein quantification was performed using the Bradford method. Lysates (100 µg) were separated on 4–12% Bis-Tris gradient gels and were transferred to PVDF membranes. Membranes were washed and incubated with antibodies directed against LIN28B (Cell Signaling Technology, sc-1916) and actin (Santa Cruz Biotechnology, sc-1616).

**LIN28B knockdown and monitoring of cell growth.**

For routine maintenance, cells were grown in RPMI 1640 complete medium (Gibco, 22400) containing 10% FBS (Hyclone, SH 30073-03), 1× antibiotic antimycotic (Gibco, 15240-062) and 2 mM l-glutamine (Gibco, 20030). On day 0, cells were seeded in triplicate into antibiotic-free medium in 96-well RT-CES plates (ACEA). On day 1, using Dharmafect (Dharmacon, T-2001-02), cells were transiently transfected with 50 nM of either a non-targeting negative control siRNA (Dharmacon, D-00810-10-05) or an siRNA directed against LIN28B (L-028584-01-0005). Real-time cell growth was monitored every 30 min for at least 96 h using the RT-CES system, as previously described. Data presented are representative of at least three independent experiments. To monitor efficiency of LIN28B knockdown, transfecion was performed as described, and 48 h later, RNA was isolated using the Qiagen mini extraction kit. Total RNA (200 ng) was primed with oligo(dT) and reverse transcribed using Superscript II reverse transcriptase (Invitrogen). Quantitative RT-PCR using TaqMan gene expression assays (ABI) was performed. To monitor LIN28B protein knockdown, protein was isolated 72 h after transfection. Protein blotting analysis was performed using antibodies directed against LIN28B and actin.

**RT-PCR in primary neuroblastomas.**

TaqMan Gene expression assays for HACE1 (Hs00410879_m1) and LIN28B (Hs01013729_m1) were purchased through Applied Biosystems. Reactions were set up in duplicate, using 10 ng of cDNA in a 10-µl reaction that contained 200 nM of probe, 900 nM of each amplification primer and 1× Real-time PCR Master Mix (Applied Biosystems). Standard curves were generated using serial dilutions of the neuroblastoma cell line Kelly. Samples were amplified on an Applied Biosystems 7900HT Sequence Detection System using standard cycling conditions, and data were collected and analyzed with SDS 2.3 software.

**Survival analysis.**

Survival analyses were performed using the methods of Kaplan and Meier. For overall survival, time was defined as the time from diagnosis until the time of death from disease or until the time of last contact if death did not occur. Individuals who were alive were censored at the time they were last known to be alive. Log-rank P values < 0.05 were considered significant.

38. Brodeur, G.M. et al. Revisions of the international criteria for neuroblastoma diagnosis, staging, and response to treatment. J. Clin. Oncol. 11, 1466–1477 (1993).
39. Shimada, H. et al. The International Neuroblastoma Pathology Classification (Shimada) System. Cancer 86, 364–372 (1999).
40. Mathew, P. et al. Detection of MYCN gene amplification in neuroblastoma by fluorescence in situ hybridization: a Pediatric Oncology Group Study. Neoplasia 3, 105–109 (2001).
41. Look, A.T. et al. Clinical relevance of tumor cell ploidy and N-myc gene amplification in childhood neuroblastoma: a Pediatric Oncology Group Study. J. Clin. Oncol. 9, 581–591 (1991).
42. Gunderson, K.L., Steemers, F.J., Lee, G., Mendoza, L.G. & Chee, M.S. A genome-wide scalable SNP genotyping assay using microarray technology. Nat. Genet. 37, 549–554 (2005).
43. Steemers, F.J. et al. Whole-genome genotyping with the single-base extension assay. Nat. Methods 3, 31–33 (2006).
44. Howie, B.N., Donnelly, P. & Marchini, J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genet. 5, e1000529 (2009).
45. Howie, B., Marchini, J. & Stephens, M. Genotype imputation with thousands of genomes. G3 (Bethesda) 1, 457–470 (2011).
46. Marchini, J., Howie, B., Myers, S., McVean, G. & Donnelly, P. A new multilocus method for genome-wide association studies by imputation of genotypes. Nat. Genet. 39, 906–913 (2007).
47. Cole, K.A. et al. A functional screen identifies miR-34a as a candidate neuroblastoma tumor suppressor gene. Mol. Cancer Res. 6, 735–742 (2008).
48. Kaplan, E.L. & Meier, P. Nonparametric estimation from incomplete observations. J. Am. Stat. Assn. 53, 457–481 (1958).