ONLINE METHODS

Sample selection. Cases were defined as a subject diagnosed with neuroblastoma or ganglioneuroblastoma and registered through the Children's Oncology Group (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, disease stage by the International Neuroblastoma Staging System[37] (INSS), International Neuroblastoma Pathology Classification[38] (INPC), *MYCN* oncogene copy number[39], DNA index (ploidy)[40], registration on clinical trial(s), event-free 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-reported 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.

Genome-wide SNP genotyping. Genotyping for both discovery and replication phases was performed using the Illumina Infinium™ II HumanHap550 and Human Quad610 BeadChip according to methods detailed elsewhere[11,41,42] and summarized here. DNA samples were surveyed for quality both by optical density spectrophotometry and the pico-green assay. A total of 750 ng of blood DNA was used to genotype each sample according to the manufacturer's guidelines. On day 1, genomic DNA was amplified 1000-1500-fold. On day 2, the amplified DNA was fragmented to ~300-600 bp, precipitated and resuspended followed by hybridization onto a BeadChip. 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 bead array, the arrayed SNP locus-specific primers (attached to beads) were extended with a single hapten-labeled dideoxynucleotide in the SBE reaction. The haptens

were subsequently detected by a multi-layer 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 was loaded directly into Illumina's genotype analysis software, GenomeStudio. Once the 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 ≥ 95% were considered for inclusion in this study.

Quality control and association testing for discovery cohort. *Overlap of the HumanHap550 v1 and v3 arrays and Quad610 array.* The discovery cohort consists 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 v3 array contains 561,288 markers, and the Quad610 array contains 620,901 markers. Overall, 535,752 markers are shared by the three arrays and analyzed in this study.

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

Inferring individuals of European ancestry. Multi-Dimensional Scaling (MDS), as implemented in the PLINK software, was used for inferring population structure in the neuroblastoma data set. Comparing self-identified ancestry with the MDS-inferred ancestry confirmed the reliability of MDS to identify genetically inferred individuals of European ancestry. In total, 2,012 neuroblastoma patients of European ancestry are clustered towards the upper left

side of the triangle (red circle) compared with 11 HapMap3 populations, and defined by Principal Component 1 (*x*-axis) less than -0.02, and Principal Component 2 (*y*-axis) greater than 0.015 (**Supplementary Fig. 1**).

Detection and elimination of cryptic relatedness and duplicated genotyping. To detect cryptic relatedness and potential duplicated genotyping within our data, we have applied a two-step procedure to calculate pairwise IBD estimates between all individual case and control subjects. First, we examined the initial MDS and retained only those individuals of inferred European ancestry (see description in previous section) and with call rates greater than 95%. Second, we recalculated genome-wide IBS estimates and IBD estimates among the remaining individuals of European ancestry using the PLINK software. This two-step procedure ensures that allele frequency differences between populations do not lead to biases in IBD estimations. We applied a stringent threshold for detecting cryptic relatedness: any pairs of subjects with IBD>0.15 were flagged and one individual 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 rate less than 99% were excluded from analysis. The call rates were calculated by the 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) less than 5% were excluded from our analysis. The MAF are calculated by the 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*-value less than 0.001 were excluded from analysis. A total of 11,953 markers were excluded from association analysis in this step.

Matching controls. Based on genome-wide IBS estimates for all pairwise comparisons among case and control subjects, we identified two matched controls for each case individual to correct the potential effects of population structure as described previously[9].

Final counts of subjects and markers passing QC. Applying the QC measures mentioned in all the previous sections, we were left with 2,101 cases, 4,202 matching controls, and 426,697 SNPs for association analysis.

Quality control and replication in 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[12]. The replication cohort after QC consisted of a total of 365 cases and 2,491 control subjects who were genotyped by the Illumina HumanHap550 or Human610-Quad BeadChip. Significant SNPs from the discovery effort ($P < 10^{-4}$) 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 method for a further validation.

Meta-analysis. Meta-analysis was performed using the inverse-variance method within the METAL[14] 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. 43) using the world-wide 1000 Genomes Project Phase 1 interim data as reference (June 2011 release). Rather than pre-selecting a reference population, we elected to follow the approach of Howie *et al.*[44] and use a multi-population reference panel with IMPUTE2. Genotypes for markers located on chromosome 6 were extracted and mapped to hg19 using the LiftOver tool

(http://genome.ucsc.edu/cgi-bin/hgLiftOver). Multi-population haplotype data from the 1000 Genomes Phase I Interim release (June 2011) were downloaded from http://mathgen.stats.ox.ac.uk/impute/data_download_1000G_phase1_interim.html. IMPUTE2 was applied with default parameters and Ne=20000. Following imputation, SNPs with MAF <1% and/or IMPUTE2-info quality score <0.8 were removed. To correctly account 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 SNPTEST[45].

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 as per manufacturer’s specifications. Data were normalized using the average normalization method provided in GenomeStudio (Illumina). Two-sided t-tests were performed at the gene level to assess differential expression in cell lines. $P < 0.05$ was considered significant.

MicroRNA expression profiling of neuroblastoma cell lines. Sample RNAs were 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. The average signal values were normalized (rank invariant) using GenomeStudio (Illumina).

LIN28B protein detection. Neuroblastoma cell lines were grown in T75 flasks under standard cell culturing conditions. Whole cell lysates were extracted with 100 μl cell extraction buffer (Invitrogen) containing protease inhibitors (Sigma) and phenylmethyl sulphonyl fluoride, briefly sonicated and rotated for 1 h at 4°C. After a 30 min 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 transferred to PVDF membranes. Membranes were then washed and incubated with antibodies directed against LIN28B (Cell Signaling, 4196) and actin (Santa Cruz, sc-1616).

***LIN28B* knockdown and monitoring of cell growth.** For routine maintenance, cells were grown in RPMI 1640 complete media (Gibco 22400) containing 10% fetal bovine serum (Hyclone SH 30073-03), 1X antibiotic antimycotic (Gibco 15240-062) and 2 mM L-glutamine (Gibco 25030). On day 0, cells were seeded in triplicate in antibiotic-free media 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[3,46]. Data presented are representative of at least three independent experiments. To monitor efficiency of *LIN28B* RNA knockdown, transfection was performed as described above, and 48 h later, RNA was isolated using the Qiagen mini extraction kit. Two hundred ng of total RNA was oligodT primed 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. Western blotting analysis was performed using antibodies directed against LIN28B (Cell Signaling, 4196) and actin (Santa Cruz, sc-1616).

Realtime 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 which contained 200 nM concentration of probe, 900 nM of each amplification primer and 1 X of 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 collected and analyzed with SDS 2.3 software.

Survival analysis. Survival analyses were performed using the methods of Kaplan and Meier[47]. For overall-survival (OS), 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. Patients who were alive were censored at the time last known alive. Log-rank *P*-values < 0.05 were considered significant.