

# Germline 16p11.2 Microdeletion Predisposes to Neuroblastoma

Laura E. Egolf,<sup>1,2,3</sup> Zalman Vaksman,<sup>2,3,4</sup> Gonzalo Lopez,<sup>2,3,4</sup> Jo Lynne Rokita,<sup>2,3,4</sup> Apexa Modi,<sup>2,3,5</sup> Patricia V. Basta,<sup>6,7</sup> Hakon Hakonarson,<sup>8,9</sup> Andrew F. Olshan,<sup>6,7</sup> and Sharon J. Diskin<sup>1,2,3,4,5,10,\*</sup>

Neuroblastoma is a cancer of the developing sympathetic nervous system. It is diagnosed in 600–700 children per year in the United States and accounts for 12% of pediatric cancer deaths. Despite recent advances in our understanding of this malignancy's complex genetic architecture, the contribution of rare germline variants remains undefined. Here, we conducted a genome-wide analysis of large (>500 kb), rare (<1%) germline copy number variants (CNVs) in two independent, multi-ethnic cohorts totaling 5,585 children with neuroblastoma and 23,505 cancer-free control children. We identified a 550-kb deletion on chromosome 16p11.2 significantly enriched in neuroblastoma cases (0.39% of cases and 0.03% of controls;  $p = 3.34 \times 10^{-9}$ ). Notably, this CNV corresponds to a known microdeletion syndrome that affects approximately one in 3,000 children and confers risk for diverse developmental phenotypes including autism spectrum disorder and other neurodevelopmental disorders. The CNV had a substantial impact on neuroblastoma risk, with an odds ratio of 13.9 (95% confidence interval = 5.8–33.4). The association remained significant when we restricted our analysis to individuals of European ancestry in order to mitigate potential confounding by population stratification (0.42% of cases and 0.03% of controls;  $p = 4.10 \times 10^{-8}$ ). We used whole-genome sequencing (WGS) to validate the deletion in paired germline and tumor DNA from 12 cases. Finally, WGS of four parent-child trios revealed that the deletion primarily arose *de novo* without maternal or paternal bias. This finding expands the clinical phenotypes associated with 16p11.2 microdeletion syndrome to include cancer, and it suggests that disruption of the 16p11.2 region may dysregulate neurodevelopmental pathways that influence both neurological phenotypes and neuroblastoma.

Current knowledge of predisposition to neuroblastoma (MIM: 256700) is incomplete. Approximately 1%–2% of cases occur in the context of familial disease with a dominant mode of inheritance and are largely explained by germline mutations in the anaplastic lymphoma kinase (ALK) gene<sup>1</sup> (MIM: 105590) or the paired like homeobox 2B (PHOX2B) gene<sup>2,3</sup> (MIM: 603851). However, the vast majority of neuroblastomas arise sporadically without family history. Genome-wide association studies (GWAS) have identified common variants that confer risk for sporadic neuroblastoma at over a dozen genetic loci,<sup>4–13</sup> including a common copy number variant (CNV) at 1q21.1 (MIM: 613017).<sup>14</sup> Notably, several of the genes identified through this approach have been functionally validated as neuroblastoma oncogenes or tumor suppressors that influence both tumor initiation and maintenance.<sup>15–20</sup> However, the common variants identified to date have only low or moderate effects on neuroblastoma risk. Higher-impact variants are expected to be rarer due to purifying selection,<sup>21</sup> but the role of rare variants in sporadic neuroblastoma has not been extensively explored. Recent sequencing efforts have reported rare germline single-nucleotide variants (SNVs) and indels which affect cancer-associated genes in children with neuroblastoma,<sup>22–24</sup>

but power to detect true disease associations has been limited due to the lack of sufficient sample sizes available. In contrast, large CNVs can be reliably detected using high-density single-nucleotide polymorphism (SNP) arrays, and hence CNV-based studies often do not suffer the same limitations. Rare CNVs have recently been implicated in multiple complex diseases, including cancer.<sup>25–28</sup> Here, we hypothesized that rare germline CNVs contribute substantially to neuroblastoma risk with larger effect sizes than do common variants previously identified through neuroblastoma GWAS.

To identify rare CNVs associated with neuroblastoma, we analyzed a large cohort of 5,585 neuroblastoma cases and 23,505 cancer-free control children SNP-genotyped at the Children's Hospital of Philadelphia (CHOP). All neuroblastoma specimens were obtained with informed consent at original diagnosis through Children's Oncology Group (COG) member institutions. DNA from peripheral blood lymphocytes or bone marrow was provided for genotyping. Neuroblastoma cases were not selected for tumor stage or other characteristics, so a range of clinical presentations across low-, intermediate-, and high-risk neuroblastoma were represented (Table 1). The control group included cancer-free children recruited after informed

<sup>1</sup>Cell and Molecular Biology Graduate Group, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA; <sup>2</sup>Division of Oncology, Department of Pediatrics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA; <sup>3</sup>Center for Childhood Cancer Research, Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA; <sup>4</sup>Department of Biomedical and Health Informatics, Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA; <sup>5</sup>Genomics and Computational Biology Graduate Group, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA; <sup>6</sup>Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina-Chapel Hill, Chapel Hill, NC 27599, USA; <sup>7</sup>Lineberger Comprehensive Cancer Center, School of Medicine, University of North Carolina-Chapel Hill, Chapel Hill, NC 27599, USA; <sup>8</sup>Division of Human Genetics, Department of Pediatrics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA; <sup>9</sup>Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA; <sup>10</sup>Abramson Family Cancer Research Institute, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA

\*Correspondence: [diskin@email.chop.edu](mailto:diskin@email.chop.edu)  
<https://doi.org/10.1016/j.ajhg.2019.07.020>.

© 2019 American Society of Human Genetics.



**Table 1. Summary of Clinical Covariates for Neuroblastoma Cases before and after Quality Control**

Clinical Covariate	Discovery Cohort		Replication Cohort	
	Before Quality Control (n = 3,688)	After Quality Control (n = 3,309)	Before Quality Control (n = 2,384)	After Quality Control (n = 2,276)
<b>Sex</b>				
female	1,668 (45.2%)	1,496 (45.2%)	1,149 (48.2%)	1,093 (45.6%)
male	1,967 (53.3%)	1,777 (53.7%)	1,232 (51.7%)	1,180 (49.2%)
unavailable	53 (1.4%)	36 (1.1%)	4 (0.1%)	3 (0.1%)
<b>Age at Diagnosis</b>				
<18 months	1,705 (46.2%)	1,553 (46.9%)	1,030 (43.2%)	978 (40.8%)
≥ 18 months	1,930 (52.3%)	1,720 (52.0%)	1,351 (56.7%)	1,295 (54.0%)
unavailable	53 (1.4%)	36 (1.1%)	4 (0.1%)	3 (0.1%)
<b>INSS<sup>a</sup> Stage</b>				
stage 1	645 (17.5%)	592 (17.9%)	402 (16.9%)	391 (16.3%)
stage 2	514 (13.9%)	473 (14.3%)	364 (15.3%)	339 (14.1%)
stage 3	581 (15.8%)	542 (16.4%)	380 (15.9%)	362 (15.1%)
stage 4	1,644 (44.6%)	1,435 (43.4%)	1,105 (46.4%)	1,063 (44.4%)
stage 4S	228 (6.2%)	210 (6.3%)	130 (5.5%)	118 (4.9%)
unavailable	76 (2.1%)	57 (1.7%)	4 (0.1%)	3 (0.1%)
<b>Risk</b>				
low	1,195 (32.4%)	1,095 (33.1%)	682 (28.6%)	651 (27.2%)
intermediate	753 (20.4%)	700 (21.2%)	567 (23.8%)	537 (22.4%)
high	1,596 (43.3%)	1,398 (42.2%)	1,034 (43.4%)	995 (41.5%)
unavailable	144 (3.9%)	116 (3.5%)	102 (4.2%)	93 (3.9%)
<b>MYCN Status</b>				
amplified	631 (17.1%)	551 (16.7%)	425 (17.8%)	411 (17.2%)
non-amplified	2,746 (74.5%)	2,481 (75.0%)	1,768 (74.2%)	1,682 (70.2%)
unavailable	315 (8.4%)	277 (8.4%)	192 (8.0%)	183 (7.6%)

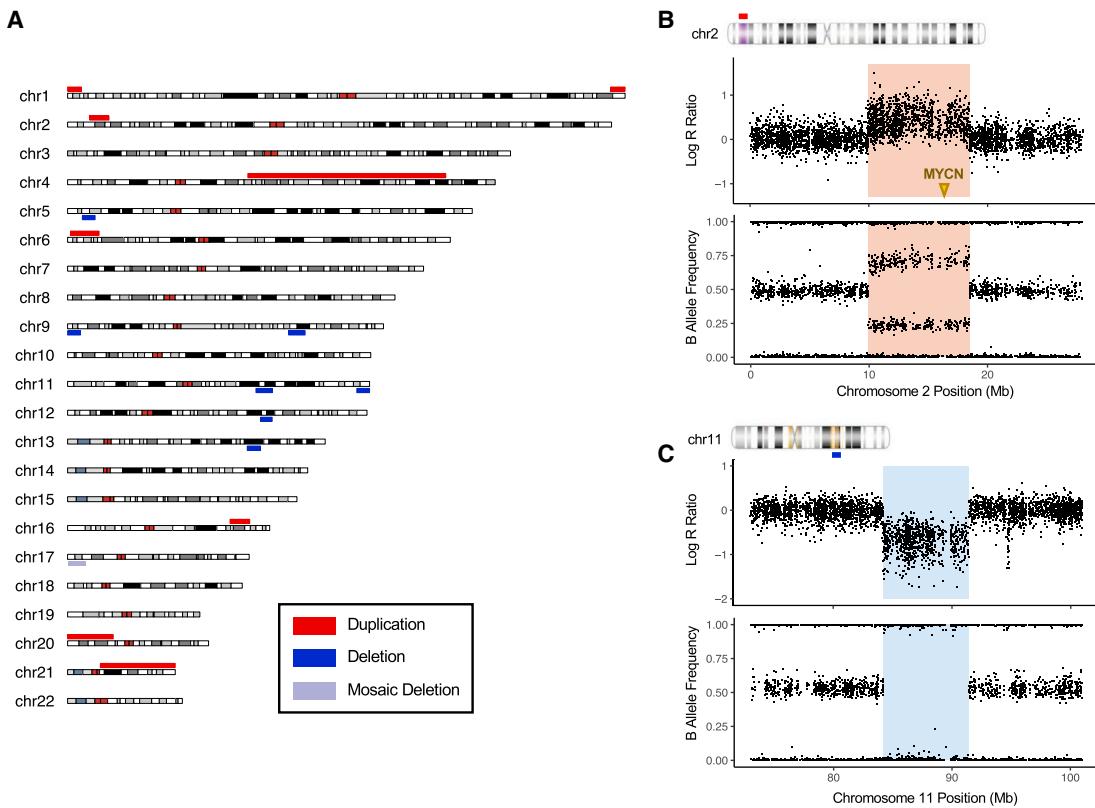
<sup>a</sup>INSS = International Neuroblastoma Staging System.

consent through the CHOP Health Care Network by the Center for Applied Genomics (CAG) and genotyped together with the neuroblastoma cases on matched genotyping platforms at the CAG. The CHOP Institutional Review Board was responsible for oversight of this study. The diverse case and control datasets included individuals of European, African, East Asian, South Asian, and Hispanic descent (Figure S1).

The initial dataset before quality control included 6,072 neuroblastoma cases and 24,242 cancer-free controls. We divided these samples into discovery and replication cohorts based on SNP array platform: samples genotyped on the Illumina HumanHap550 and Human610-Quad arrays were used for discovery (n = 3,688 cases and 9,229 controls) and samples genotyped more recently on the HumanOmniExpress arrays served as an independent replication cohort (n = 2,384 cases and 15,013 controls, Table S1). We restricted the analysis to high-quality SNPs

(see Supplemental Material and Methods in [Supplemental Data](#)) shared across all array types within each cohort (402,743 and 506,045 SNPs for discovery and replication, respectively). CNV calls were then generated using a circular binary segmentation algorithm implemented in Nexus Copy Number (BioDiscovery) with adjustment for waves created by GC content variability (Table S2). We removed excessively noisy samples and those with contamination from circulating tumor cells and DNA (see Supplemental Material and Methods in [Supplemental Data](#)), resulting in a final discovery cohort of 3,309 cases and 8,855 controls and a replication cohort of 2,276 cases and 14,650 controls (Table S1).

We first examined megabase-scale chromosomal abnormalities in our combined cohort of 5,585 neuroblastoma cases. Chromosomal abnormalities of this size are detectable by conventional cytogenetic analysis as well as by SNP array and comparative genomic hybridization array



**Figure 1. Megabase-Scale Germline Chromosomal Abnormalities Including 2p Gain and 11q Loss Are Observed in 0.27% of Neuroblastoma Cases**

(A) Germline deletions and duplications larger than 5 Mb on autosomal chromosomes are plotted in a summary karyotype (karyoploteR<sup>74</sup>) showing the entire cohort of 5,585 neuroblastoma cases. We observed 16 events affecting 15 individuals (the 1q44 terminal duplication and mosaic 17p13 terminal deletion were present in the same individual).

(B) Log R ratio and B allele frequency plots for the 8.6-Mb germline duplication of 2p25.1-p24.2 (red shading) in a high-risk, MYCN-amplified case. The location of the MYCN oncogene is indicated by an arrow.

(C) Log R ratio and B allele frequency plots for the 7.2-Mb 11q14 interstitial deletion (blue shading) in an intermediate-risk, MYCN-non-amplified case.

(array CGH) technologies. Several abnormalities have been identified in children with neuroblastoma and published as case reports and case series over the last 50 years,<sup>29</sup> but their overall frequency has not been established. Here, we report that deletions and duplications larger than 5 Mb on autosomal chromosomes were present in 0.27% of individuals in our large unselected neuroblastoma cohort (Figure 1A and Figure S3). Notably, a 2p25.1-p24.2 duplication encompassing the oncogene MYCN (MIM: 164840) and an interstitial 11q14 deletion were observed in neuroblastoma cases but not in 23,505 cancer-free controls (Figures 1B and 1C). Additionally, a terminal deletion on 11q24.3-q25 was observed in one case and one control. MYCN amplification and 11q loss are frequent somatic alterations in neuroblastoma and are predictive of poor outcome.<sup>30,31</sup> Reports of germline 2p gain and 11q deletion are sparse in the literature, but several of the reported cases have presented with neuroblastoma.<sup>29,32-34</sup> The megabase-scale events reported here passed our quality control (see Supplemental Material and Methods in Supplemental Data) and showed allelic ratios consistent with a germline deletion or duplication. However, we

cannot completely rule out the possibility of contamination from circulating tumor DNA, particularly for the 2p25.1-p24.2 gain, because matched tumor data were not available. Finally, trisomy 21 (MIM: 190685) was observed in only one out of 5,585 neuroblastoma cases in this cohort, which is low relative to the general population prevalence of approximately one out of 1,200 individuals.<sup>35</sup> This is consistent with previous studies and suggests a decreased incidence of Down Syndrome in neuroblastoma cases.<sup>29,36</sup>

To narrow our focus to a small number of rare, potentially high-impact CNVs for further analyses and to limit the interference of artifacts, we considered only CNVs that were longer than 500 kb and that met stringent filtering criteria (Table S2.). As expected, large CNVs were rare. We identified 4,668 high-confidence large CNVs across the entire sample set, which came to 0.160 per individual on average. Large duplications were more common than deletions (0.114 and 0.046 per individual, respectively). The overall burden of large CNVs was similar in neuroblastoma cases and cancer-free controls. We observed an average of 0.154 and 0.157 CNVs per

**Table 2. Association of 16p11.2 Microdeletion with Neuroblastoma**

Cohort	Total Cases	Cases with 16p11.2 Deletion	Total Controls	Controls with 16p11.2 Deletion	Deletion Frequency in Cases	Deletion Frequency in Controls	p Value	Odds Ratio (95% CI) <sup>a</sup>
<b>All Subjects</b>								
discovery	3,309	12	8,855	2	0.36%	0.02%	8.28x10 <sup>-6</sup>	16.1 (3.6–148.2)
replication	2,276	10	14,650	5 <sup>b</sup>	0.44%	0.03%	2.99x10 <sup>-6</sup>	12.9 (4.0–48.2)
meta-analysis	-	-	-	-	-	-	3.34x10 <sup>-9</sup>	13.9 (5.8–33.4)
<b>European Only</b>								
discovery	2,219	7	6,236	2	0.32%	0.03%	1.83x10 <sup>-3</sup>	9.9 (1.9–97.4)
replication	1,340	8	12,065	4 <sup>b</sup>	0.60%	0.03%	3.35x10 <sup>-6</sup>	18.1 (4.8–82.2)
meta-analysis	-	-	-	-	-	-	4.10x10 <sup>-8</sup>	14.5 (5.6–37.6)

<sup>a</sup>CI = Confidence Interval<sup>b</sup>One control possesses a mosaic 16p11.2 microdeletion.

case and control, respectively, in the discovery cohort ( $p = 0.70$ ) and 0.178 and 0.161 CNVs per case and control in the replication cohort ( $p = 0.07$ ). Similarly, no significant differences were observed between cases and controls when considering deletions and duplications separately.

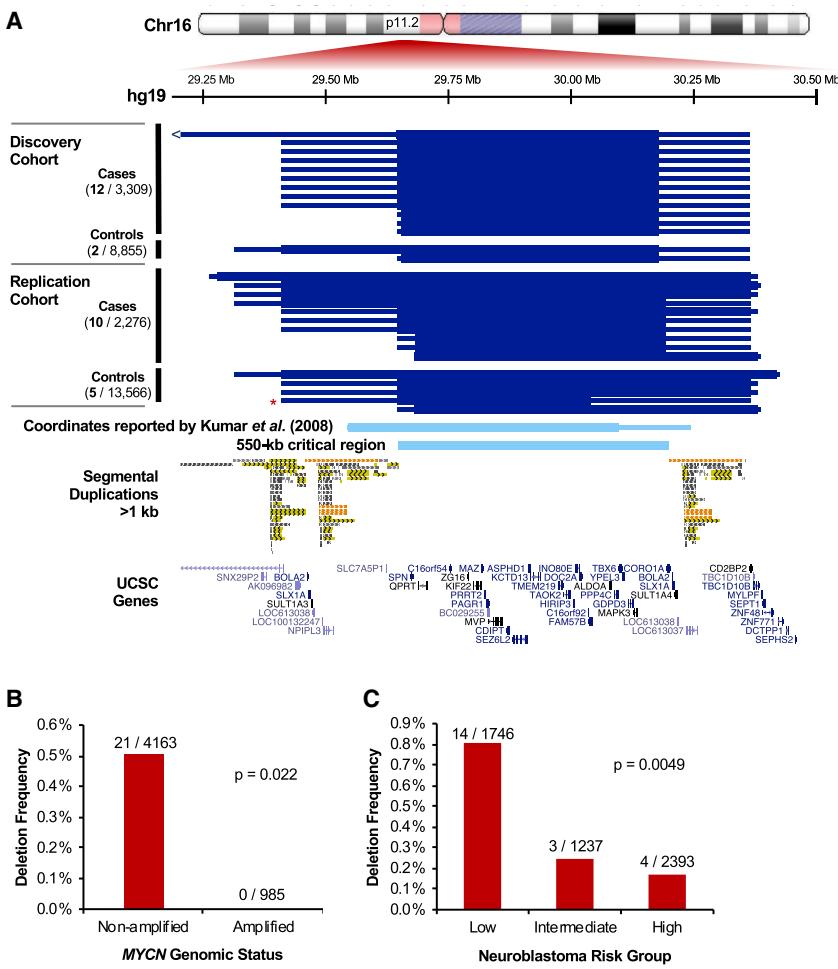
We next carried out a regional association study to identify CNVs that were larger than 500 kb and enriched in neuroblastoma in our discovery cohort of 3,309 cases and 8,855 controls. We collapsed CNVs >500 kb in size that were recurrent in three or more cases (approximately 0.1%) into copy number variable regions (CNVRs), requiring at least 100 kb minimum overlap. Consistent with the overall rarity of large CNVs, no region was deleted or duplicated in more than 0.63% of neuroblastoma cases or 0.58% of controls in the discovery cohort. We observed 40 recurrent CNVRs: seven deletions and 33 duplications (Table S3). We tested these recurrent CNVRs for association with neuroblastoma using Fisher's exact test, applying a Bonferroni-corrected significance threshold ( $p = 0.00125$ ). With these restrictions, the only CNVR reaching genome-wide significance was a 550 kb single-copy deletion on chromosome 16p11.2 (in 0.36% of cases and 0.02% of controls;  $p = 8.3 \times 10^{-6}$ ; Table 2 and Figure 2A). We validated this association in the independent replication cohort ( $p = 3.0 \times 10^{-6}$ ). Meta-analysis detected no heterogeneity between the discovery and replication cohorts ( $p = 0.8$ ) and yielded a combined  $p$  value of  $3.34 \times 10^{-9}$  and an odds ratio of 13.9 (95% confidence interval: 5.8–33.4). This represents a considerably larger impact on neuroblastoma susceptibility than was seen with common variants previously identified through GWAS, for which odds ratios range from 1.2 to 3.0.<sup>37</sup>

Notably, deletion of this 550 kb region on 16p11.2 is a recognized genetic syndrome (MIM: 611913) that confers increased risk for autism spectrum disorder, developmental delay, intellectual disability, seizures, macrocephaly, early-onset obesity, increased body mass index, and birth defects.<sup>38–45</sup> Chromosome 16p is rich in segmental duplications, which serve as hotspots for non-allelic homologous

recombination (NAHR) during meiosis or mitosis and give rise to reciprocal microdeletions and microduplications in several regions. Five recurrent breakpoint regions have been defined.<sup>44</sup> The deletion we observed in neuroblastoma is referred to as the proximal (or typical) deletion and is characterized by breakpoint (BPs) in the segmental duplications near 29.6 and 30.2 Mb (BP4–BP5). The 550 kb unique segment between the segmental duplications (29.65–30.20 Mb) is considered the critical region, and this segment contains 25 annotated protein-coding genes. We did not observe any distal (BP1–BP3 or BP2–BP3) deletions in our cohort of 5,585 neuroblastoma cases, and we observed proximal and distal 16p11.2 microduplications only at frequencies consistent with their population prevalence (two and three cases, respectively).

These 16p11.2 microdeletions were observed in 0.39% of neuroblastoma cases and 0.03% of controls overall. The frequency observed in our control cohort is consistent with previous estimates of 16p11.2 deletion prevalence, which range from 0.01%–0.04% in adult populations<sup>38,43,45,46</sup> and 0.03% in a screen of French-Canadian newborns.<sup>47</sup> We observed deletions in all ethnic groups except South Asians, likely due to the small number of South Asian individuals included in the study (Figure S2). The association of 16p11.2 microdeletions with neuroblastoma remained significant in the discovery and replication cohorts when considering European subjects only (0.42% of cases and 0.03% of controls;  $p = 4.10 \times 10^{-8}$ ; odds ratio = 14.5; 95% confidence interval = 5.6–37.6; Table 2).

We next tested the 16p11.2 microdeletion for association with clinical, biological, and genetic covariates in our neuroblastoma cohort (Table S4). Strikingly, no cases with 16p11.2 microdeletion had somatic amplification of the oncogene MYCN ( $p = 0.022$ , Figure 2B), an event that is observed in approximately 20% of neuroblastoma tumors and associates with poor prognosis. The deletion was more frequent in cases classified as low- or intermediate-risk according to the COG risk stratification method



**Figure 2. 16p11.2 Microdeletion Associates with Neuroblastoma and Is Enriched in MYCN Non-Amplified, Low-Risk Cases**

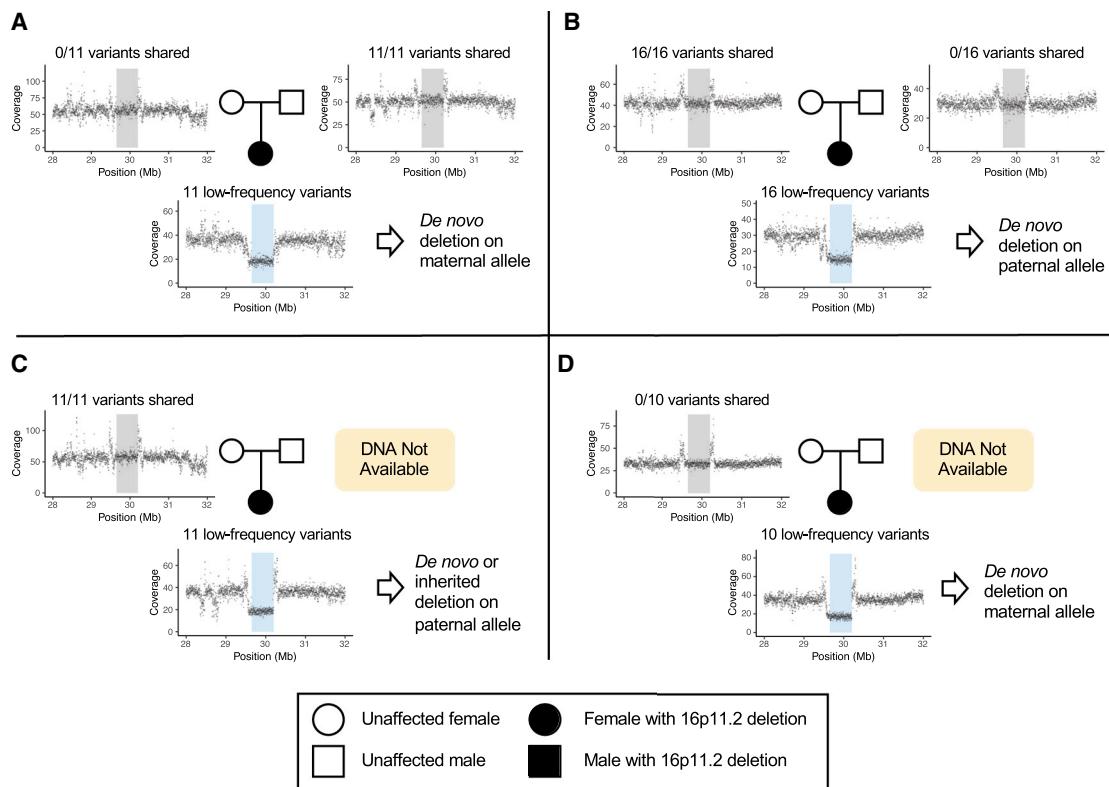
(A) 16p11.2 microdeletions identified in neuroblastoma cases and cancer-free controls in each analysis cohort (discovery and replication) are plotted in UCSC Genome Browser.<sup>75</sup> Minimum and maximum deletion coordinates approximated by SNP array are represented by the thick and thin blue bars, respectively. The red asterisk denotes one mosaic deletion identified in a control (the maximum end coordinate for this sample was extended to encompass a region of decreased probe intensity not called by the segmentation algorithm). The deletion coordinates reported by Kumar et al.<sup>39</sup> are shown for reference (converted from hg18 to hg19 using UCSC liftOver<sup>76</sup>). These coordinates match the 16p11.2 proximal CNV, a 550-kb critical region flanked by segmental duplications. This unique critical region (29.65–30.20 Mb) contains 25 annotated protein-coding genes.

(B–C) 16p11.2 microdeletion was tested for association with clinical covariates in 21 cases with available annotation using Fisher's exact test. The deletion associated significantly ( $p < 0.05$ ) with MYCN amplification status (B) and Children's Oncology Group risk classification (C).

( $p = 0.0049$ , Figure 2C), even after adjusting for MYCN amplification ( $p = 0.046$ ). It also trended toward association with tumor stage as described by the International Neuroblastoma Staging System ( $p = 0.080$ ), and there was a slight enrichment for Stage 1 and 2 disease. There was no significant association with sex ( $p = 1$ ) or age at diagnosis ( $p = 0.68$ ). We compared survival rates between low-risk cases with and without the deletion and did not detect a significant difference in overall or event-free survival at current sample sizes ( $p = 0.54$  and  $p = 0.23$ , respectively; Figure S4). Finally, we restricted our analyses to European individuals and tested 16p11.2 microdeletion for association with genotypes at eight known neuroblastoma susceptibility loci previously identified through GWAS. We did not detect any significant differences in allele frequencies after multiple testing corrections (Table S5). The enrichment of MYCN non-amplified, low-risk cases among those with 16p11.2 microdeletion may suggest that the deletion predisposes individuals to neuroblastoma through a mechanism that is more likely to give rise to low-risk disease. Some neuroblastoma-associated GWAS variants have also shown enrichment in distinct disease subsets such as high- or low-risk, MYCN amplified, or 11q-deleted cases.<sup>5,10,13,48</sup>

To validate these 16p11.2 microdeletions and gain further genetic insight, we performed 30 $\times$  whole-genome sequencing (WGS) on germline and tumor DNA from seven neuroblastoma cases with the microdeletion. We also utilized WGS data for five additional microdeletion-carrying neuroblastoma cases that were sequenced through the Therapeutically Applicable Research to Generate Effective Treatments (TARGET) program or the Gabriella Miller Kids First (GMKF) program. Germline deletions were confirmed in all 12 individuals, and the deletions were also detected in all 11 available matched tumors (Figure S5). As estimated by SNP array, all deletions covered the 550 kb critical region from 29.65–30.20 Mb and had breakpoints falling within the flanking segmental duplications.

We next used the WGS data for 11 tumors from 16p11.2 microdeletion cases to profile genome-wide somatic copy number changes (Figure S6). Segmental chromosomal alterations, which clinically correlate with high-risk disease,<sup>49</sup> were observed in two individuals. Nine tumors harbored no large copy number changes or only whole-chromosome gains without segmental alterations, consistent with low-risk classification. Chromosome 16 was gained in three hyperploid low-risk tumors, but did not harbor any other large alterations. We also assessed the 16p11.2 region for focal copy number changes in these tumors (Figure S5). In most tumors, read coverage within the 16p11.2 deleted region was decreased by approximately



**Figure 3. 16p11.2 Microdeletion Arises *de novo* in Neuroblastoma without Allelic Bias**

Coverage is plotted in 2-kb bins for four neuroblastoma cases with parental DNA available: two parent-child trios (A and B) and two mother-child duos (C and D). All parents show normal coverage within the 16p11.2 critical region (20.65–30.20 Mb, gray shading), whereas the children show a 50% decrease in coverage; this result confirms 16p11.2 microdeletion (blue shading). The number of low-frequency variants (<5% in 1000 Genomes) found within each child's single remaining allele is displayed above the child's coverage plot. The fraction of these variants the child shares with each parent is displayed above the parent's plot.

50% relative to the surrounding region; this result was similar to the decrease in relative coverage seen with the germline one-copy deletion. However, two tumors had slightly lower relative coverage (37%–38%), and one had higher relative coverage (65%), indicating that the region may have undergone additional somatic NAHR-mediated rearrangements in these three tumors.

We next hypothesized that damaging germline or somatic variants in the unaffected allele might exacerbate the effect of the deletion by disrupting one or more of the 25 protein-coding genes within the deleted region. We identified five rare (allele frequency < 1%), predicted-damaging (CADD score  $\geq 20$ ) germline SNVs and indels affecting these protein-coding genes among the 12 cases with germline WGS data available (Table S6), but we found no somatic SNVs or indels meeting these criteria in 11 matched tumors. We identified two germline missense variants of potential clinical relevance by using the Human Gene Mutation Database: The R386H variant in *SEZ6L2* is weakly associated with autism spectrum disorder (ASD [MIM: 209850]),<sup>50</sup> and the P216L variant in *PRRT2* has been reported in Landau-Kleffner syndrome (MIM: 245570), a rare childhood neurological disorder.<sup>51</sup> It is unclear whether these and other variants within 16p11.2 play a functional role in neuroblastoma initiation or progression, but follow-up

studies on these protein-coding variants as well as noncoding and structural variants are warranted.

Lastly, to determine whether these 16p11.2 microdeletions arose *de novo* or were inherited, we considered four cases for which parental DNA had been collected through the Neuroblastoma Epidemiology in North America (NENA) study<sup>52</sup> and WGS had been performed through the GMKF Pediatric Research Program (Figure 3 and Table S7). By comparing relative coverage and low-frequency variants within the deleted region in the children to the same data from their parents, we concluded that the deletion arose *de novo* in three of four children; of these, two deletions arose on the maternal allele and one on the paternal allele. The fourth deletion was either inherited from the father or arose *de novo* on the paternal allele. This predominantly *de novo* inheritance pattern was also observed for ASD-associated 16p11.2 microdeletions, where 90% of deletion cases were *de novo* or mosaic in the germline.<sup>53</sup> However, in ASD, *de novo* 16p11.2 deletions exhibited a strong maternal bias: 89% arose on the maternal haplotype.<sup>53</sup> In contrast, here we detected 16p11.2 deletions occurring *de novo* on both the maternal and paternal alleles. Maternal ages ranged from 17–32 and paternal ages from 18–42, showing no apparent trend with *de novo* deletion origin (Table S7).

16p11.2 microdeletion is associated with a diverse array of phenotypes including ASD, developmental delay, seizures, macrocephaly, and obesity.<sup>35–42</sup> The association of 16p11.2 microdeletion with neuroblastoma further expands the breadth of the deletion's clinical manifestations to include cancer. Notably, this region has previously been implicated in the regulation of proliferation and growth: while 16p11.2 microdeletion associates with macrocephaly and increased body weight, the reciprocal 16p11.2 microduplication associates with microcephaly and decreased body weight.<sup>40,43</sup> Similar reciprocal deletion/duplication growth phenotypes are also observed with the neurodevelopment-associated CNVs at 1q21.1 (MIM: 612474, 612475)<sup>54</sup> and 22q11.2 (MIM: 192430, 608363).<sup>55</sup> In general, neurodevelopmental disorders are enriched for mutations in cancer-associated genes which control proliferation and differentiation.<sup>56</sup> This supports the notion that genetic variants such as 16p11.2 microdeletion could predispose individuals to both cancer and neurodevelopmental phenotypes by perturbing these cellular functions. For 16p11.2 microdeletion, this effect may partially be attributed to dysregulation of the MAPK/ERK pathway caused by deletion of *MAPK3* (MIM: 601795), encoding ERK1, and *MVP* (MIM: 605088). Deletion of a region syntenic to 16p11.2 causes elevated ERK activity in neural precursors in mice<sup>57,58</sup> but the impact on other lineages, such as those that give rise to neuroblastoma, is unknown. Aberrant MAPK signaling contributes to many cancers, including neuroblastoma, where an enrichment of mutations in the MAPK pathway has been observed at relapse.<sup>59</sup>

The association of 16p11.2 microdeletion with neuroblastoma is especially noteworthy given this cancer's neurodevelopmental origin. Neuroblastoma is thought to arise from improper differentiation of neural crest cells of the sympathetic adrenal lineage, giving rise to primary tumors in sympathetic nervous system tissues. The pleiotropic effects of 16p11.2 microdeletion on neuroblastoma and neurodevelopmental phenotypes such as ASD may be explained by dysregulation of developmental programs involved in both central and peripheral nervous system development. Most studies on 16p11.2 microdeletion syndrome have focused on central nervous system defects,<sup>60–67</sup> so additional research on the peripheral nervous system is needed. It is possible that aberrant neural crest development plays an underappreciated role in the pathology of 16p11.2 microdeletion syndrome. The neural crest has important functions in development of the brain and the adrenal and endocrine systems, facial bones, heart, and other tissues. Dysregulation of these systems could partially explain phenotypes of 16p11.2 microdeletion including language and learning impairment, macrocephaly, dysmorphic facial features, and cardiac malformation.<sup>68</sup> This possibility is supported by a recent study which suggests that altered neural crest activity explains some of these features in the context of ASD.<sup>69</sup> Additionally, several

16p11.2 genes were implicated in neural crest-related phenotypes in a zebrafish loss-of-function screen.<sup>70</sup> Future research on the role of the 16p11.2 region in neural crest development could uncover novel biology and shed light on the clinical manifestations of the 16p11.2 CNV.

The phenotypes exhibited by 16p11.2 microdeletion carriers can vary dramatically, even within the same family. Some carriers have no identifiable phenotypic abnormalities, while others are severely affected with multiple deficits.<sup>68</sup> This makes the deletion syndrome extremely challenging to diagnose and manage, and it suggests that other factors cooperate to influence pathogenesis. Both common and rare variants have been shown to modify neurodevelopmental outcomes and other phenotypes in 16p11.2 microdeletion carriers.<sup>71–73</sup> Given that most deletion carriers do not develop clinically detectable neuroblastoma, 16p11.2 deletion alone is likely not sufficient for neuroblastoma tumorigenesis. Additional genetic, epigenetic, and environmental factors probably contributed to initiation and maintenance in individuals harboring the microdeletion. We identified five rare, predicted-damaging variants in protein-coding genes within the 16p11.2 region that could potentially function as second hits, but the significance of these variants and other genetic and epigenetic alterations requires further investigation.

Currently, complete medical histories are not available for the individuals with neuroblastoma profiled in this study. An important future direction for this work is to re-contact families of children with 16p11.2 microdeletion and define the co-occurrence of neuroblastoma with other traits typical of the microdeletion syndrome. This will help determine whether counseling or surveillance for neuroblastoma might be appropriate for children diagnosed with 16p11.2 microdeletion, or conversely, whether germline testing for children newly diagnosed with neuroblastoma should include 16p11.2 microdeletion so that the other phenotypes associated with the syndrome can be monitored.

Overall, this study identifies a rare germline CNV that substantially impacts neuroblastoma risk, highlighting the potential clinical relevance of rare variants in this often deadly pediatric cancer. This finding expands the already diverse clinical outcomes associated with 16p11.2 microdeletion syndrome to include cancer, and it raises questions regarding the role of the 16p11.2 region in neural crest development.

## Data Availability

Data analyzed in this paper are available through the database of Genotypes and Phenotypes (dbGaP). SNP array data are available through accession phs000124.v3.p1. WGS data generated through TARGET are available through accession phs000218.v21.p7. WGS data generated through GMKF are available through dbGaP accession phs001436.v1.p1 and through the Kids First Data Resource

Portal. Additional WGS data generated for this study have been deposited to dbGaP: phs000124.v3.p1.

## Accession Numbers

The dbGaP accession number for the WGS data newly generated for this study and reported in this paper is phs000124.v3.p1.

## Supplemental Data

Supplemental Data can be found online at <https://doi.org/10.1016/j.ajhg.2019.07.020>.

## Acknowledgments

This work was supported by National Institutes of Health (NIH) awards R01CA204974, R03CA230366, R01CA132887, R01CA124709, X01HL136997, and U2CHL138346. L.E.E. was supported in part by NIH grant T32GM008216, and A.M. was supported in part by T32HG000046, also from the NIH. Neuroblastoma case DNA used for SNP array genotyping and for the WGS generated for this manuscript was collected and bio-banked through the Children's Oncology Group (COG). The platform-matched control genotyping data were generated and supplied by the Center for Applied Genomics (CAG) at CHOP. DNA extraction from Neuroblastoma Epidemiology in North America (NENA) blood or buccal samples was carried out in the UNC BioSpecimen Processing Core. Whole-genome sequencing from Gabriella Miller Kids First (GMKF) was delivered on the CAVATICA analysis platform by the Kids First Pediatric Data Resource Center.

## Declaration of Interests

The authors declare no competing interests.

Received: April 17, 2019

Accepted: July 29, 2019

Published: August 29, 2019

## Web Resources

ANNOVAR version 2016Feb01, [annovar.openbioinformatics.org/](http://annovar.openbioinformatics.org/)  
dbGaP, <https://www.ncbi.nlm.nih.gov/gap>  
Gabriella Miller Kids First (GMKF) Pediatric Research Program,  
<https://commonfund.nih.gov/kidsfirst/>  
GATK version 4.0.2.0, <https://software.broadinstitute.org/gatk/>  
GenomeStudio version 2.0 (Illumina), [https://support.illumina.com/array/array\\_software/genomestudio/downloads.html](https://support.illumina.com/array/array_software/genomestudio/downloads.html)  
karyoploteR version 1.8.8, <http://www.bioconductor.org/packages/karyoploteR>  
Kids First Data Resource Portal, <https://kidsfirstdrc.org/>  
Nexus Copy Number versions 8.0 and 10.0 (BioDiscovery),  
<https://www.biodescovery.com/products/Nexus-Copy-Number>  
OMIM, <http://www.omim.org/>  
PLINK version 1.9, <https://www.cog-genomics.org/plink2>  
R version 3.5.2, <https://www.r-project.org/>  
SnpEff version 4.3, <http://snpeff.sourceforge.net/>  
Therapeutically Applicable Research to Generate Effective Treatments (TARGET) Program, <https://ocg.cancer.gov/programs/target>  
UCSC Genome Browser, <http://genome.ucsc.edu/>

UCSC liftOver, <https://genome.ucsc.edu/cgi-bin/hgLiftOver>  
VarScan2 version 2.3.9, <http://varscan.sourceforge.net/>

## References

1. Mossé, Y.P., Laudenslager, M., Longo, L., Cole, K.A., Wood, A., Attiyeh, E.F., Laquaglia, M.J., Sennett, R., Lynch, J.E., Perri, P., et al. (2008). Identification of ALK as a major familial neuroblastoma predisposition gene. *Nature* **455**, 930–935.
2. Mosse, Y.P., Laudenslager, M., Khazi, D., Carlisle, A.J., Winter, C.L., Rappaport, E., and Maris, J.M. (2004). Germline PHOX2B mutation in hereditary neuroblastoma. *Am. J. Hum. Genet.* **75**, 727–730.
3. Trochet, D., Bourdeaut, F., Janoueix-Lerosey, I., Deville, A., de Pontual, L., Schleiermacher, G., Coze, C., Philip, N., Frébourg, T., Munnich, A., et al. (2004). Germline mutations of the paired-like homeobox 2B (PHOX2B) gene in neuroblastoma. *Am. J. Hum. Genet.* **74**, 761–764.
4. Maris, J.M., Mosse, Y.P., Bradfield, J.P., Hou, C., Monni, S., Scott, R.H., Asgharzadeh, S., Attiyeh, E.F., Diskin, S.J., Laudenslager, M., et al. (2008). Chromosome 6p22 locus associated with clinically aggressive neuroblastoma. *N. Engl. J. Med.* **358**, 2585–2593.
5. Capasso, M., Devoto, M., Hou, C., Asgharzadeh, S., Glessner, J.T., Attiyeh, E.F., Mosse, Y.P., Kim, C., Diskin, S.J., Cole, K.A., et al. (2009). Common variations in BARD1 influence susceptibility to high-risk neuroblastoma. *Nat. Genet.* **41**, 718–723.
6. Wang, K., Diskin, S.J., Zhang, H., Attiyeh, E.F., Winter, C., Hou, C., Schnepf, R.W., Diamond, M., Bosse, K., Mayes, P.A., et al. (2011). Integrative genomics identifies LMO1 as a neuroblastoma oncogene. *Nature* **469**, 216–220.
7. Diskin, S.J., Capasso, M., Schnepf, R.W., Cole, K.A., Attiyeh, E.F., Hou, C., Diamond, M., Carpenter, E.L., Winter, C., Lee, H., et al. (2012). Common variation at 6q16 within HACE1 and LIN28B influences susceptibility to neuroblastoma. *Nat. Genet.* **44**, 1126–1130.
8. McDaniel, L.D., Conkrite, K.L., Chang, X., Capasso, M., Vaksman, Z., Oldridge, D.A., Zachariou, A., Horn, M., Diamond, M., Hou, C., et al. (2017). Common variants upstream of MLF1 at 3q25 and within CPZ at 4p16 associated with neuroblastoma. *PLoS Genet.* **13**, e1006787.
9. Capasso, M., Diskin, S.J., Totaro, F., Longo, L., De Mariano, M., Russo, R., Cimmino, F., Hakonarson, H., Tonini, G.P., Devoto, M., et al. (2013). Replication of GWAS-identified neuroblastoma risk loci strengthens the role of BARD1 and affirms the cumulative effect of genetic variations on disease susceptibility. *Carcinogenesis* **34**, 605–611.
10. Nguyen, B., Diskin, S.J., Capasso, M., Wang, K., Diamond, M.A., Glessner, J., Kim, C., Attiyeh, E.F., Mosse, Y.P., Cole, K., et al. (2011). Phenotype restricted genome-wide association study using a gene-centric approach identifies three low-risk neuroblastoma susceptibility Loci. *PLoS Genet.* **7**, e1002026.
11. Latorre, V., Diskin, S.J., Diamond, M.A., Zhang, H., Hakonarson, H., Maris, J.M., and Devoto, M. (2012). Replication of neuroblastoma SNP association at the BARD1 locus in African-Americans. *Cancer Epidemiol. Biomarkers Prev.* **21**, 658–663.
12. Capasso, M., McDaniel, L.D., Cimmino, F., Cirino, A., Formicola, D., Russell, M.R., Raman, P., Cole, K.A., and Diskin, S.J. (2017). The functional variant rs34330 of CDKN1B is associated with risk of neuroblastoma. *J. Cell. Mol. Med.* **21**, 3224–3230.

13. Chang, X., Zhao, Y., Hou, C., Glessner, J., McDaniel, L., Diamond, M.A., Thomas, K., Li, J., Wei, Z., Liu, Y., et al. (2017). Common variants in MMP20 at 11q22.2 predispose to 11q deletion and neuroblastoma risk. *Nat. Commun.* **8**, 569.
14. Diskin, S.J., Hou, C., Glessner, J.T., Attiyeh, E.F., Laudenslager, M., Bosse, K., Cole, K., Mossé, Y.P., Wood, A., Lynch, J.E., et al. (2009). Copy number variation at 1q21.1 associated with neuroblastoma. *Nature* **459**, 987–991.
15. Russell, M.R., Penikis, A., Oldridge, D.A., Alvarez-Dominguez, J.R., McDaniel, L., Diamond, M., Padovan, O., Raman, P., Li, Y., Wei, J.S., et al. (2015). CASC15-S is a tumor suppressor lncRNA at the 6p22 neuroblastoma susceptibility locus. *Cancer Res.* **75**, 3155–3166.
16. Pandey, G.K., Mitra, S., Subhash, S., Hertwig, F., Kanduri, M., Mishra, K., Fransson, S., Ganeshram, A., Mondal, T., Bandaru, S., et al. (2014). The risk-associated long noncoding RNA NBAT-1 controls neuroblastoma progression by regulating cell proliferation and neuronal differentiation. *Cancer Cell* **26**, 722–737.
17. Bosse, K.R., Diskin, S.J., Cole, K.A., Wood, A.C., Schnepf, R.W., Norris, G., Nguyen, B., Jagannathan, J., Laquaglia, M., Winter, C., et al. (2012). Common variation at BARD1 results in the expression of an oncogenic isoform that influences neuroblastoma susceptibility and oncogenicity. *Cancer Res.* **72**, 2068–2078.
18. Zhu, S., Zhang, X., Weichert-Leahy, N., Dong, Z., Zhang, C., Lopez, G., Tao, T., He, S., Wood, A.C., Oldridge, D., et al. (2017). LMO1 Synergizes with MYCN to Promote Neuroblastoma Initiation and Metastasis. *Cancer Cell* **32**, 310–323.e5.
19. Molenaar, J.J., Domingo-Fernández, R., Ebus, M.E., Lindner, S., Koster, J., Drabek, K., Mestdagh, P., van Sluis, P., Valentijn, L.J., van Nes, J., et al. (2012). LIN28B induces neuroblastoma and enhances MYCN levels via let-7 suppression. *Nat. Genet.* **44**, 1199–1206.
20. Schnepf, R.W., Khurana, P., Attiyeh, E.F., Raman, P., Chodosh, S.E., Oldridge, D.A., Gagliardi, M.E., Conkrite, K.L., Asgharzadeh, S., Seeger, R.C., et al. (2015). A LIN28B-RAN-AURKA Signaling Network Promotes Neuroblastoma Tumorigenesis. *Cancer Cell* **28**, 599–609.
21. Lek, M., Karczewski, K.J., Minikel, E.V., Samocha, K.E., Banks, E., Fennell, T., O'Donnell-Luria, A.H., Ware, J.S., Hill, A.J., Cummings, B.B., et al.; Exome Aggregation Consortium (2016). Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285–291.
22. Pugh, T.J., Morozova, O., Attiyeh, E.F., Asgharzadeh, S., Wei, J.S., Auclair, D., Carter, S.L., Cibulskis, K., Hanna, M., Kiezun, A., et al. (2013). The genetic landscape of high-risk neuroblastoma. *Nat. Genet.* **45**, 279–284.
23. Parsons, D.W., Roy, A., Yang, Y., Wang, T., Scollon, S., Bergstrom, K., Kerstein, R.A., Gutierrez, S., Petersen, A.K., Bavle, A., et al. (2016). Diagnostic yield of clinical tumor and germline whole-exome sequencing for children with solid tumors. *JAMA Oncol.* **2**, 616–624.
24. Zhang, J., Walsh, M.F., Wu, G., Edmonson, M.N., Gruber, T.A., Easton, J., Hedges, D., Ma, X., Zhou, X., Yergeau, D.A., et al. (2015). Germline Mutations in Predisposition Genes in Pediatric Cancer. *N. Engl. J. Med.* **373**, 2336–2346.
25. Stefansson, H., Rujescu, D., Cichon, S., Pietiläinen, O.P.H., Ingason, A., Steinberg, S., Fossdal, R., Sigurdsson, E., Sigmundsson, T., Buizer-Voskamp, J.E., et al.; GROUP (2008). Large recurrent microdeletions associated with schizophrenia. *Nature* **455**, 232–236.
26. Pinto, D., Pagnamenta, A.T., Klei, L., Anney, R., Merico, D., Regan, R., Conroy, J., Magalhaes, T.R., Correia, C., Abrahams, B.S., et al. (2010). Functional impact of global rare copy number variation in autism spectrum disorders. *Nature* **466**, 368–372.
27. Park, R.W., Kim, T.M., Kasif, S., and Park, P.J. (2015). Identification of rare germline copy number variations over-represented in five human cancer types. *Mol. Cancer* **14**, 1–12.
28. Marshall, C.R., Howrigan, D.P., Merico, D., Thiruvahindrapuram, B., Wu, W., Greer, D.S., Antaki, D., Shetty, A., Hollmans, P.A., Pinto, D., et al.; Psychosis Endophenotypes International Consortium; and CNV and Schizophrenia Working Groups of the Psychiatric Genomics Consortium (2017). Contribution of copy number variants to schizophrenia from a genome-wide study of 41,321 subjects. *Nat. Genet.* **49**, 27–35.
29. Satgé, D., Moore, S.W., Stiller, C.A., Niggli, F.K., Pritchard-Jones, K., Bown, N., Bénard, J., and Plantaz, D. (2003). Abnormal constitutional karyotypes in patients with neuroblastoma: a report of four new cases and review of 47 others in the literature. *Cancer Genet. Cytogenet.* **147**, 89–98.
30. Attiyeh, E.F., London, W.B., Mossé, Y.P., Wang, Q., Winter, C., Khazi, D., McGrady, P.W., Seeger, R.C., Look, A.T., Shimada, H., et al.; Children's Oncology Group (2005). Chromosome 1p and 11q deletions and outcome in neuroblastoma. *N. Engl. J. Med.* **353**, 2243–2253.
31. Thompson, D., Vo, K.T., London, W.B., Fischer, M., Ambros, P.F., Nakagawara, A., Brodeur, G.M., Matthay, K.K., and DuBois, S.G. (2016). Identification of patient subgroups with markedly disparate rates of MYCN amplification in neuroblastoma: A report from the International Neuroblastoma Risk Group project. *Cancer* **122**, 935–945.
32. Mosse, Y., Greshock, J., King, A., Khazi, D., Weber, B.L., and Maris, J.M. (2003). Identification and high-resolution mapping of a constitutional 11q deletion in an infant with multifocal neuroblastoma. *Lancet Oncol.* **4**, 769–771.
33. Passariello, A., De Brasi, D., Defferrari, R., Genesio, R., Tufano, M., Mazzocco, K., Capasso, M., Migliorati, R., Martinsson, T., Siani, P., et al. (2013). Constitutional 11q14-q22 chromosome deletion syndrome in a child with neuroblastoma MYCN single copy. *Eur. J. Med. Genet.* **56**, 626–634.
34. Lurie, I.W. (2014). Clinical manifestations of partial trisomy 2p. *Cytogenet. Genome Res.* **144**, 28–30.
35. Presson, A.P., Partyka, G., Jensen, K.M., Devine, O.J., Rasmussen, S.A., McCabe, L.L., and McCabe, E.R.B. (2013). Current estimate of Down Syndrome population prevalence in the United States. *J. Pediatr.* **163**, 1163–1168.
36. Satgé, D., Sasco, A.J., Carlsen, N.L., Stiller, C.A., Rubie, H., Hero, B., de Bernardi, B., de Kraker, J., Coze, C., Kogner, P., et al. (1998). A lack of neuroblastoma in Down syndrome: a study from 11 European countries. *Cancer Res.* **58**, 448–452.
37. Ritenour, L.E., Randall, M.P., Bosse, K.R., and Diskin, S.J. (2018). Genetic susceptibility to neuroblastoma: current knowledge and future directions. *Cell Tissue Res.* **372**, 287–307.
38. Weiss, L.A., Shen, Y., Korn, J.M., Arking, D.E., Miller, D.T., Fossdal, R., Saemundsen, E., Stefansson, H., Ferreira, M.A.R., Green, T., et al.; Autism Consortium (2008). Association between microdeletion and microduplication at 16p11.2 and autism. *N. Engl. J. Med.* **358**, 667–675.
39. Kumar, R.A., KaraMohamed, S., Sudi, J., Conrad, D.F., Brune, C., Badner, J.A., Gilliam, T.C., Nowak, N.J., Cook, E.H., Jr.,

- Dobyns, W.B., and Christian, S.L. (2008). Recurrent 16p11.2 microdeletions in autism. *Hum. Mol. Genet.* **17**, 628–638.
40. Shinawi, M., Liu, P., Kang, S.H.L., Shen, J., Belmont, J.W., Scott, D.A., Probst, F.J., Craigen, W.J., Graham, B.H., Pursley, A., et al. (2010). Recurrent reciprocal 16p11.2 rearrangements associated with global developmental delay, behavioural problems, dysmorphism, epilepsy, and abnormal head size. *J. Med. Genet.* **47**, 332–341.
  41. Rosenfeld, J.A., Coppinger, J., Bejjani, B.A., Girirajan, S., Eichler, E.E., Shaffer, L.G., and Ballif, B.C. (2010). Speech delays and behavioral problems are the predominant features in individuals with developmental delays and 16p11.2 microdeletions and microduplications. *J. Neurodev. Disord.* **2**, 26–38.
  42. Walters, R.G., Jacquemont, S., Valsesia, A., de Smith, A.J., Martinet, D., Andersson, J., Falchi, M., Chen, F., Andrieux, J., Lobbens, S., et al. (2010). A new highly penetrant form of obesity due to deletions on chromosome 16p11.2. *Nature* **463**, 671–675.
  43. Jacquemont, S., Reymond, A., Zufferey, F., Harewood, L., Walters, R.G., Kutalik, Z., Martinet, D., Shen, Y., Valsesia, A., Beckmann, N.D., et al. (2011). Mirror extreme BMI phenotypes associated with gene dosage at the chromosome 16p11.2 locus. *Nature* **478**, 97–102.
  44. Zufferey, F., Sherr, E.H., Beckmann, N.D., Hanson, E., Maillard, A.M., Hippolyte, L., Macé, A., Ferrari, C., Kutalik, Z., Andrieux, J., et al.; Simons VIP Consortium; and 16p11.2 European Consortium (2012). A 600 kb deletion syndrome at 16p11.2 leads to energy imbalance and neuropsychiatric disorders. *J. Med. Genet.* **49**, 660–668.
  45. Macé, A., Tuke, M.A., Deelen, P., Kristiansson, K., Mattsson, H., Nöukas, M., Sapkota, Y., Schick, U., Porcu, E., Rüeger, S., et al. (2017). CNV-association meta-analysis in 191,161 European adults reveals new loci associated with anthropometric traits. *Nat. Commun.* **8**, 1–11.
  46. Malhotra, D., and Sebat, J. (2012). CNVs: harbingers of a rare variant revolution in psychiatric genetics. *Cell* **148**, 1223–1241.
  47. Tucker, T., Giroux, S., Clément, V., Langlois, S., Friedman, J.M., and Rousseau, F. (2013). Prevalence of selected genomic deletions and duplications in a French-Canadian population-based sample of newborns. *Mol. Genet. Genomic Med.* **1**, 87–97.
  48. Hungate, E.A., Applebaum, M.A., Skol, A.D., Vaksman, Z., Diamond, M., McDaniel, L., Volchenboum, S.L., Stranger, B.E., Maris, J.M., Diskin, S.J., et al. (2017). Evaluation of Genetic Predisposition for MYCN-Amplified Neuroblastoma. *J. Natl. Cancer Inst.* **109**, 1–4.
  49. Schleiermacher, G., Janoueix-Lerosey, I., Ribeiro, A., Klijnienko, J., Couturier, J., Pierron, G., Mosseri, V., Valent, A., Auger, N., Plantaz, D., et al. (2010). Accumulation of segmental alterations determines progression in neuroblastoma. *J. Clin. Oncol.* **28**, 3122–3130.
  50. Kumar, R.A., Marshall, C.R., Badner, J.A., Babatz, T.D., Mukamel, Z., Aldinger, K.A., Sudi, J., Brune, C.W., Goh, G., Karamahamed, S., et al. (2009). Association and mutation analyses of 16p11.2 autism candidate genes. *PLoS ONE* **4**, e4582.
  51. Conroy, J., McGgettigan, P.A., McCreary, D., Shah, N., Collins, K., Parry-Fielder, B., Moran, M., Hanrahan, D., Deonna, T.W., Korff, C.M., et al. (2014). Towards the identification of a genetic basis for Landau-Kleffner syndrome. *Epilepsia* **55**, 858–865.
  52. Mazul, A.L., Siega-Riz, A.M., Weinberg, C.R., Engel, S.M., Zou, F., Carrier, K.S., Basta, P.V., Vaksman, Z., Maris, J.M., Diskin, S.J., et al. (2016). A family-based study of gene variants and maternal folate and choline in neuroblastoma: a report from the Children's Oncology Group. *Cancer Causes Control* **27**, 1209–1218.
  53. Duyzend, M.H., Nuttle, X., Coe, B.P., Baker, C., Nickerson, D.A., Bernier, R., and Eichler, E.E. (2016). Maternal Modifiers and Parent-of-Origin Bias of the Autism-Associated 16p11.2 CNV. *Am. J. Hum. Genet.* **98**, 45–57.
  54. Mefford, H.C., Sharp, A.J., Baker, C., Itsara, A., Jiang, Z., Buysse, K., Huang, S., Maloney, V.K., Crolla, J.A., Baralle, D., et al. (2008). Recurrent rearrangements of chromosome 1q21.1 and variable pediatric phenotypes. *N. Engl. J. Med.* **359**, 1685–1699.
  55. Rump, P., de Leeuw, N., van Essen, A.J., Verschueren-Bemelmans, C.C., Veenstra-Knol, H.E., Swinkels, M.E., Oostdijk, W., Ruivenkamp, C., Reardon, W., de Munnik, S., et al. (2014). Central 22q11.2 deletions. *Am. J. Med. Genet. A* **164A**, 2707–2723.
  56. Ernst, C. (2016). Proliferation and Differentiation Deficits are a Major Convergence Point for Neurodevelopmental Disorders. *Trends Neurosci.* **39**, 290–299.
  57. Pucilowska, J., Vithayathil, J., Tavares, E.J., Kelly, C., Karlo, J.C., and Landreth, G.E. (2015). The 16p11.2 deletion mouse model of autism exhibits altered cortical progenitor proliferation and brain cytoarchitecture linked to the ERK MAPK pathway. *J. Neurosci.* **35**, 3190–3200.
  58. Pucilowska, J., Vithayathil, J., Pagani, M., Kelly, C., Karlo, J.C., Robol, C., Morella, I., Gozzi, A., Brambilla, R., and Landreth, G.E. (2018). Pharmacological Inhibition of ERK Signaling Rescues Pathophysiology and Behavioral Phenotype Associated with 16p11.2 Chromosomal Deletion in Mice. *J. Neurosci.* **38**, 6640–6652.
  59. Eleveld, T.F., Oldridge, D.A., Bernard, V., Koster, J., Colmet Daage, L., Diskin, S.J., Schild, L., Bentahar, N.B., Bellini, A., Chicard, M., et al. (2015). Relapsed neuroblastomas show frequent RAS-MAPK pathway mutations. *Nat. Genet.* **47**, 864–871.
  60. Escamilla, C.O., Filanova, I., Walker, A.K., Xuan, Z.X., Holehonnur, R., Espinosa, F., Liu, S., Thyme, S.B., López-García, I.A., Mendoza, D.B., et al. (2017). Kctd13 deletion reduces synaptic transmission via increased RhoA. *Nature* **551**, 227–231.
  61. Park, S.M., Park, H.R., and Lee, J.H. (2017). MAPK3 at the Autism-Linked Human 16p11.2 Locus Influences Precise Synaptic Target Selection at *Drosophila* Larval Neuromuscular Junctions. *Mol. Cells* **40**, 151–161.
  62. Haslinger, D., Waltes, R., Yousaf, A., Lindlar, S., Schneider, I., Lim, C.K., Tsai, M.M., Garvalov, B.K., Acker-Palmer, A., Kreuzdorn, N., et al. (2018). Loss of the Chr16p11.2 ASD candidate gene QPRT leads to aberrant neuronal differentiation in the SH-SY5Y neuronal cell model. *Mol. Autism* **9**, 1–17.
  63. Richter, M., Murtaza, N., Scharrenberg, R., White, S.H., Johanns, O., Walker, S., Yuen, R.K.C., Schwanke, B., Bedürftig, B., Henis, M., et al. (2018). Altered TAOK2 activity causes autism-related neurodevelopmental and cognitive abnormalities through RhoA signaling. *Mol. Psychiatry*. <https://doi.org/10.1038/s41380-018-0025-5>.
  64. Golzio, C., Willer, J., Talkowski, M.E., Oh, E.C., Taniguchi, Y., Jacquemont, S., Reymond, A., Sun, M., Sawa, A., Gusella, J.F., et al. (2012). KCTD13 is a major driver of mirrored

- neuroanatomical phenotypes of the 16p11.2 copy number variant. *Nature* **485**, 363–367.
- 65. Grissom, N.M., McKee, S.E., Schoch, H., Bowman, N., Ha'vekes, R., O'Brien, W.T., Mahrt, E., Siegel, S., Commons, K., Portfors, C., et al. (2018). Male-specific deficits in natural reward learning in a mouse model of neurodevelopmental disorders. *Mol. Psychiatry* **23**, 544–555.
  - 66. Lin, G.N., Corominas, R., Lemmens, I., Yang, X., Tavernier, J., Hill, D.E., Vidal, M., Sebat, J., and Iakoucheva, L.M. (2015). Spatiotemporal 16p11.2 protein network implicates cortical late mid-fetal brain development and KCTD13-Cul3-RhoA pathway in psychiatric diseases. *Neuron* **85**, 742–754.
  - 67. Iyer, J., Singh, M.D., Jensen, M., Patel, P., Pizzo, L., Huber, E., Koerselman, H., Weiner, A.T., Lepanto, P., Vadodaria, K., et al. (2018). Pervasive genetic interactions modulate neurodevelopmental defects of the autism-associated 16p11.2 deletion in *Drosophila melanogaster*. *Nat. Commun.* **9**, 2548.
  - 68. Miller, D.T., Chung, W., Nasir, R., Shen, Y., Steinman, K.J., Wu, B.-L., and Hanson, E. (1993). 16p11.2 recurrent microdeletion. In *GeneReviews*, M.P. Adam, H.H. Ardinger, R.A. Pagon, S.E. Wallace, L.J.H. Bean, K. Stephens, and A. Amemiya, eds. (University of Washington, Seattle).
  - 69. Benítez-Burraco, A., Lattanzi, W., and Murphy, E. (2016). Language impairments in ASD resulting from a failed domestication of the human brain. *Front. Neurosci.* **10**, 373.
  - 70. Blaker-Lee, A., Gupta, S., McCammon, J.M., De Rienzo, G., and Sive, H. (2012). Zebrafish homologs of genes within 16p11.2, a genomic region associated with brain disorders, are active during brain development, and include two deletion dosage sensor genes. *Dis. Model. Mech.* **5**, 834–851.
  - 71. Pizzo, L., Jensen, M., Polyak, A., Rosenfeld, J.A., Mannik, K., Krishnan, A., McCready, E., Pichon, O., Le Caignec, C., Van Dijck, A., et al. (2019). Rare variants in the genetic background modulate cognitive and developmental phenotypes in individuals carrying disease-associated variants. *Genet. Med.* **21**, 816–825.
  - 72. Wu, N., Ming, X., Xiao, J., Wu, Z., Chen, X., Shinawi, M., Shen, Y., Yu, G., Liu, J., Xie, H., et al. (2015). TBX6 null variants and a common hypomorphic allele in congenital scoliosis. *N. Engl. J. Med.* **372**, 341–350.
  - 73. Girirajan, S., Rosenfeld, J.A., Cooper, G.M., Antonacci, F., Siswara, P., Itsara, A., Vives, L., Walsh, T., McCarthy, S.E., Baker, C., et al. (2010). A recurrent 16p12.1 microdeletion supports a two-hit model for severe developmental delay. *Nat. Genet.* **42**, 203–209.
  - 74. Gel, B., and Serra, E. (2017). karyoplotR: an R/Bioconductor package to plot customizable genomes displaying arbitrary data. *Bioinformatics* **33**, 3088–3090.
  - 75. Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M., and Haussler, D. (2002). The human genome browser at UCSC. *Genome Res.* **12**, 996–1006.
  - 76. Hinrichs, A.S., Karolchik, D., Baertsch, R., Barber, G.P., Bejerano, G., Clawson, H., Diekhans, M., Furey, T.S., Harte, R.A., Hsu, F., et al. (2006). The UCSC Genome Browser Database: update 2006. *Nucleic Acids Res.* **34**, D590–D598.

**Supplemental Data**

**Germline 16p11.2 Microdeletion Predisposes  
to Neuroblastoma**

**Laura E. Egolf, Zalman Vaksman, Gonzalo Lopez, Jo Lynne Rokita, Apexa Modi, Patricia  
V. Basta, Hakon Hakonarson, Andrew F. Olshan, and Sharon J. Diskin**

## **Supplemental Data**

### **SUPPLEMENTAL FIGURES**

Figure S1. Ethnicity inference through principal component analysis.

Figure S2. Inferred ethnicities for individuals with 16p11.2 microdeletion.

Figure S3. Megabase-scale germline chromosomal abnormalities in cancer-free controls.

Figure S4. Survival among low-risk neuroblastoma cases with and without 16p11.2 microdeletion.

Figure S5. Read depth within the 16p11.2 region in germline and tumor DNA from 12 neuroblastoma cases with 16p11.2 microdeletion.

Figure S6. Whole-genome copy number profiles for tumor DNA from 11 neuroblastoma cases with 16p11.2 microdeletion.

### **SUPPLEMENTAL TABLES**

Table S1. Sample sizes before and after quality control (QC) and number of samples removed during each QC step.

Table S2. Copy number variants (CNVs) larger than 500 kb. (**See Excel file**)

Sheet 1: CNVs excluding chromosome 2p25.2-2p23.1 and filtered for coverage by allelic imbalance or loss of heterozygosity

Sheet 2: Unfiltered CNVs

Table S3. Association testing for copy number variable regions (CNVRs) recurrent in at least three cases. (**See Excel file**)

Sheet 1: Deletions

Sheet 2: Duplications

Table S4. Clinical, biological, and genetic covariates for neuroblastoma cases with 16p11.2 microdeletion. (**See Excel file**)

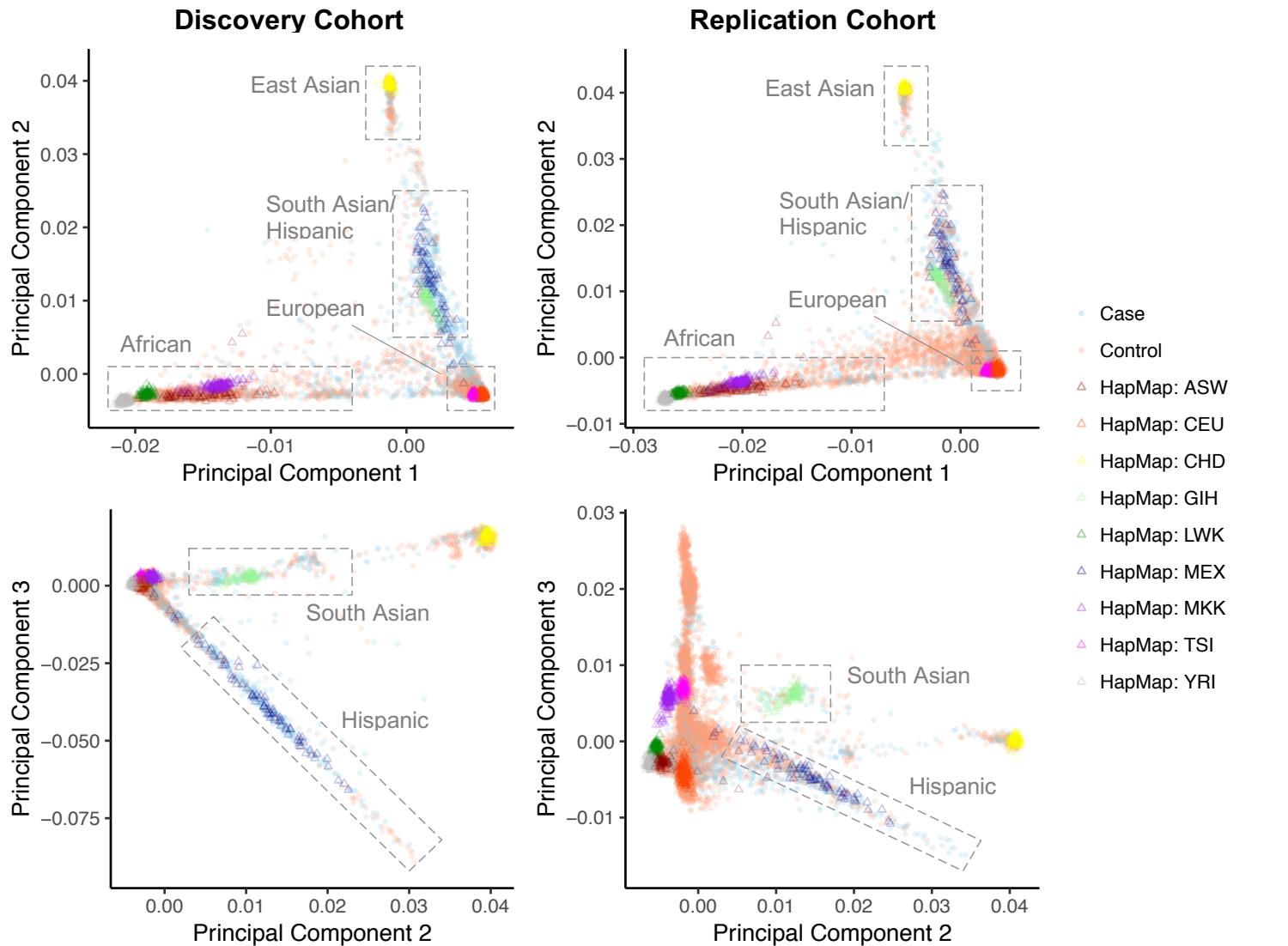
Table S5. Genotype frequencies at known neuroblastoma susceptibility loci among European neuroblastoma cases with and without 16p11.2 microdeletion.

Table S6. Rare, predicted-damaging germline variants identified in 12 neuroblastoma cases with 16p11.2 microdeletion.

Table S7. Inference of 16p11.2 deletion inheritance patterns in four families using read coverage and low-frequency variants within the deleted region. (**See Excel file**)

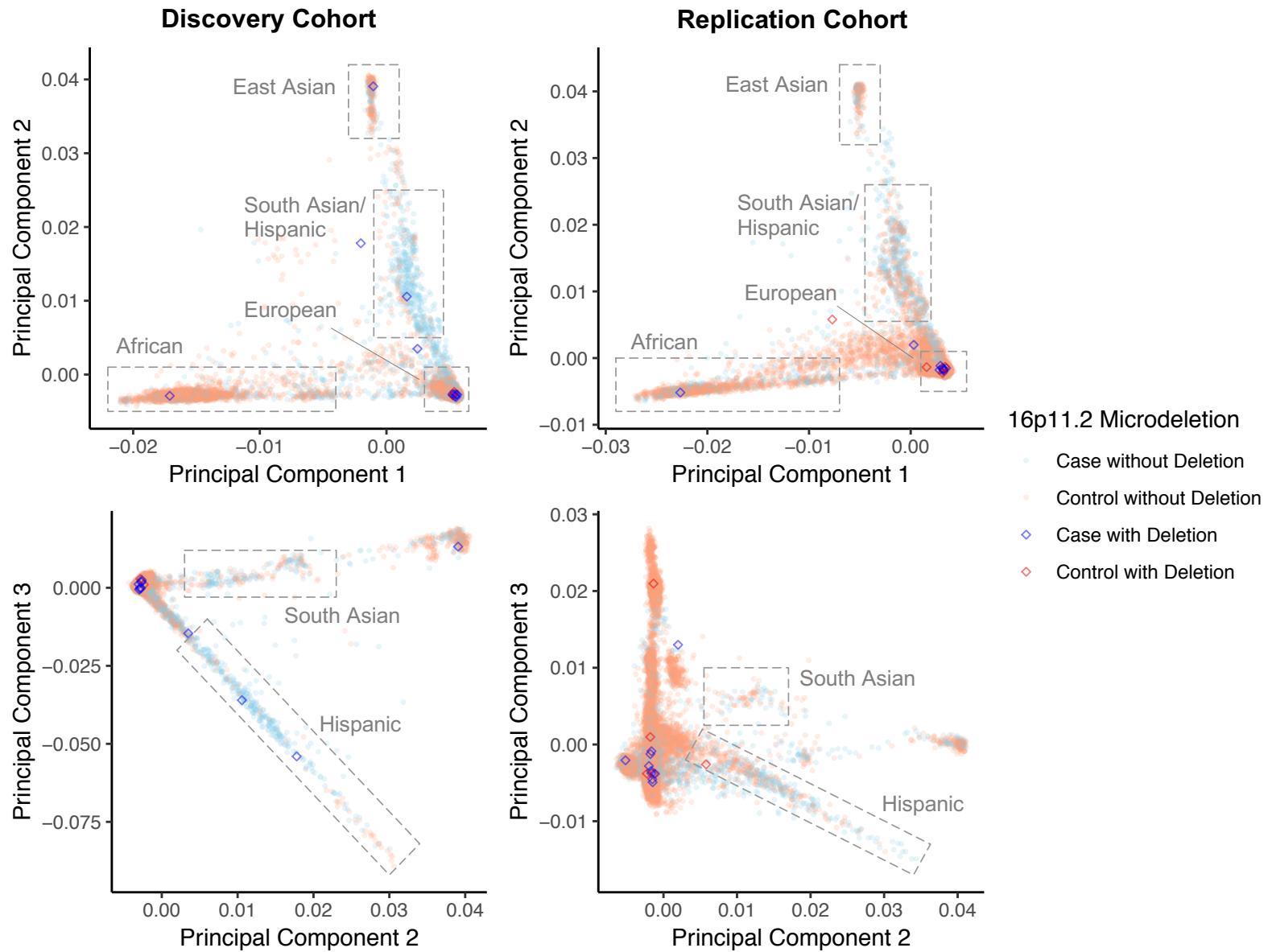
### **SUPPLEMENTAL MATERIAL AND METHODS**

### **SUPPLEMENTAL REFERENCES**



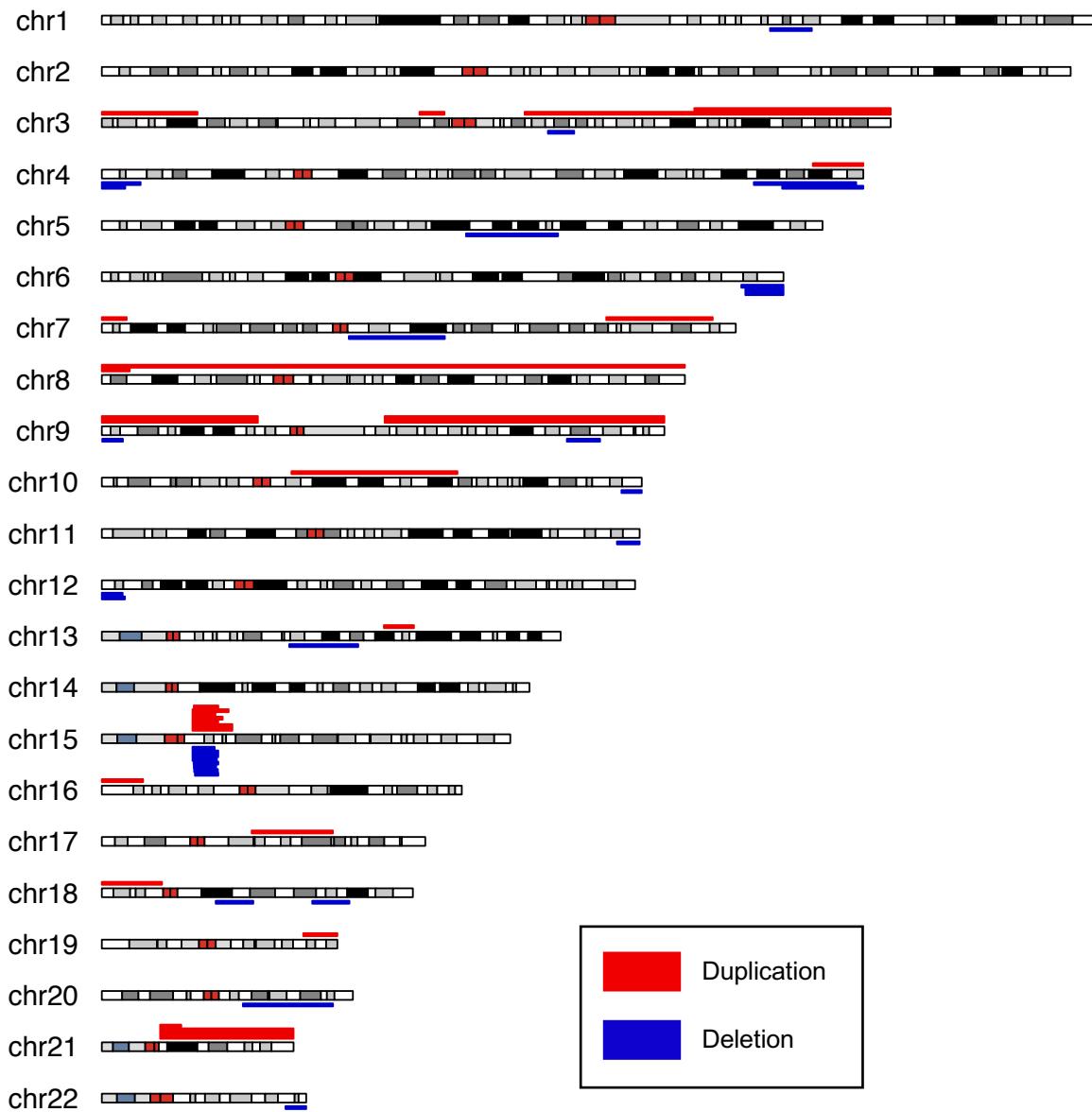
	Discovery Cohort				Replication Cohort			
	Cases		Controls		Cases		Controls	
	Count	(% of Cohort)	Count	(% of Cohort)	Count	(% of Cohort)	Count	(% of Cohort)
<b>European</b>	2219	(67.1%)	6236	(70.4%)	1340	(58.9%)	12065	(82.4%)
<b>African</b>	401	(12.1%)	2093	(23.6%)	310	(13.6%)	1060	(7.2%)
<b>Hispanic</b>	261	(7.9%)	32	(0.4%)	204	(9.0%)	222	(1.5%)
<b>East Asian</b>	70	(2.1%)	182	(2.1%)	53	(2.3%)	133	(0.9%)
<b>South Asian</b>	108	(3.3%)	57	(0.6%)	36	(1.6%)	53	(0.4%)
<b>Mixed or Unknown</b>	250	(7.6%)	255	(2.9%)	333	(14.6%)	1117	(7.6%)
<b>Total</b>	<b>3309</b>	<b>(100.0%)</b>	<b>8855</b>	<b>(100.0%)</b>	<b>2276</b>	<b>(100.0%)</b>	<b>14650</b>	<b>(100.0%)</b>

**Figure S1. Ethnicity inference through principal component analysis.** Neuroblastoma cases and cancer-free control subjects from the discovery (left) and replication (right) cohorts are plotted alongside HapMap reference populations. Individuals of European, African, and East Asian ancestry were separated from South Asian and Hispanic populations based on biplots of principal components 1 and 2. South Asian and Hispanic populations were then further differentiated using components 2 and 3. The number of individuals falling within the cutoffs defined for each inferred population, indicated by the dashed boxes, is quantified (bottom).

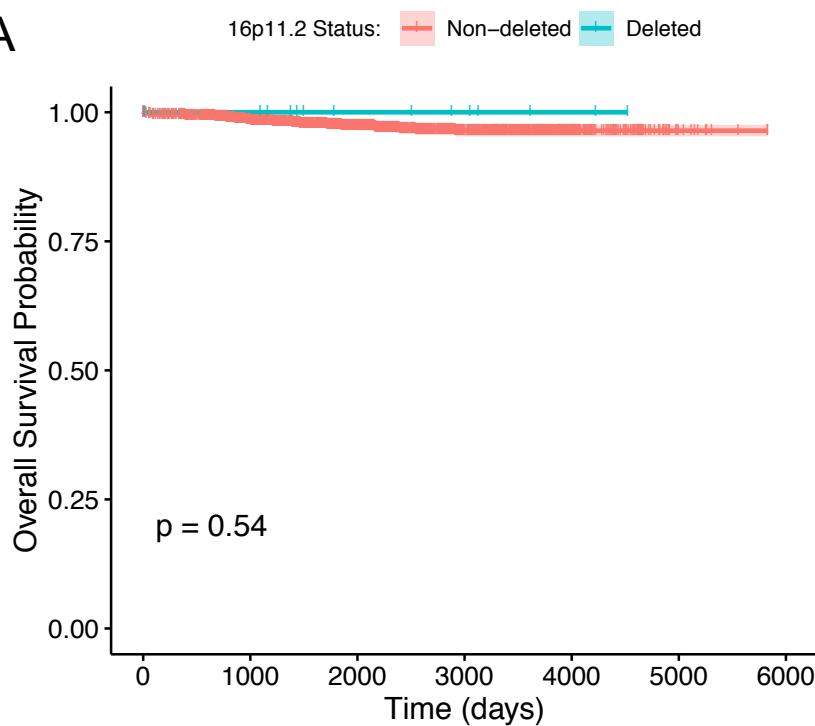
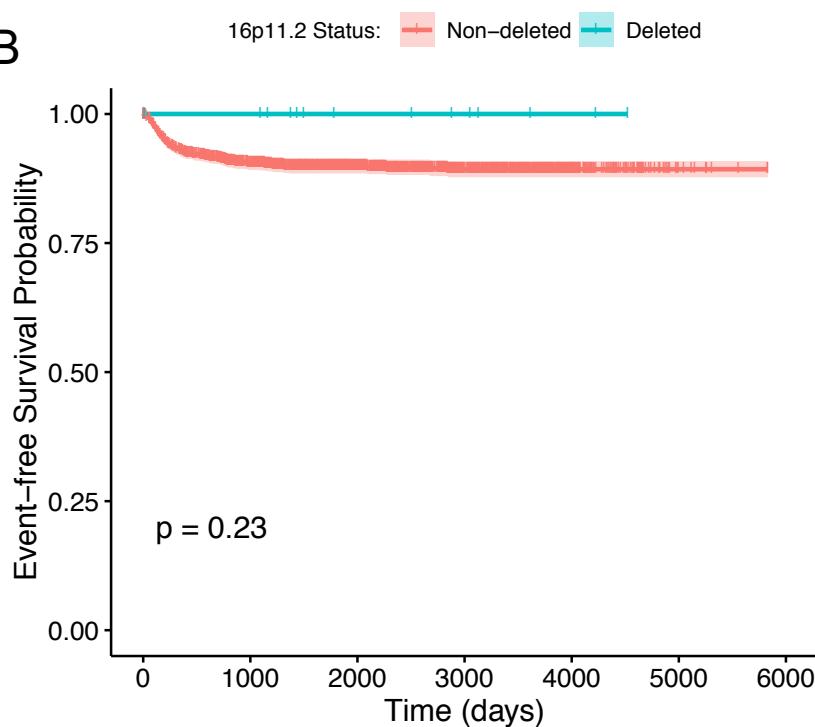


	Discovery Cohort		Replication Cohort		Combined	
	Cases	Controls	Cases	Controls	Cases	Controls
	Count (Freq.)	Count (Freq.)	Count (Freq.)	Count (Freq.)	Count (Freq.)	Count (Freq.)
<b>European</b>	7 (0.32%)	2 (0.03%)	8 (0.60%)	4 (0.03%)	15 (0.42%)	6 (0.03%)
<b>African</b>	1 (0.25%)	0 (0.00%)	1 (0.32%)	0 (0.00%)	2 (0.28%)	0 (0.00%)
<b>Hispanic</b>	1 (0.38%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	1 (0.22%)	0 (0.00%)
<b>East Asian</b>	1 (1.43%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	1 (0.81%)	0 (0.00%)
<b>South Asian</b>	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)
<b>Mixed or Unknown</b>	2 (0.80%)	0 (0.00%)	1 (0.30%)	1 (0.09%)	3 (0.51%)	1 (0.07%)
<b>Total</b>	<b>12 (0.36%)</b>	<b>2 (0.02%)</b>	<b>10 (0.44%)</b>	<b>5 (0.03%)</b>	<b>22 (0.39%)</b>	<b>7 (0.03%)</b>

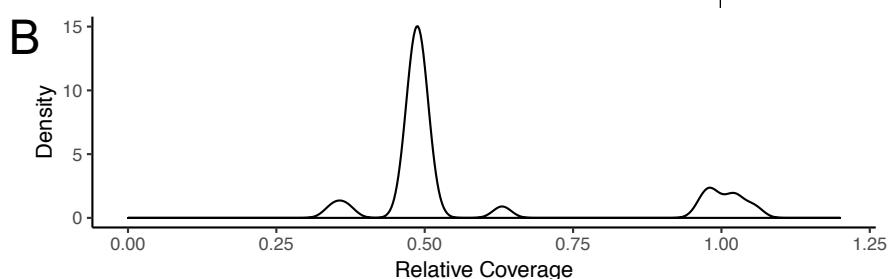
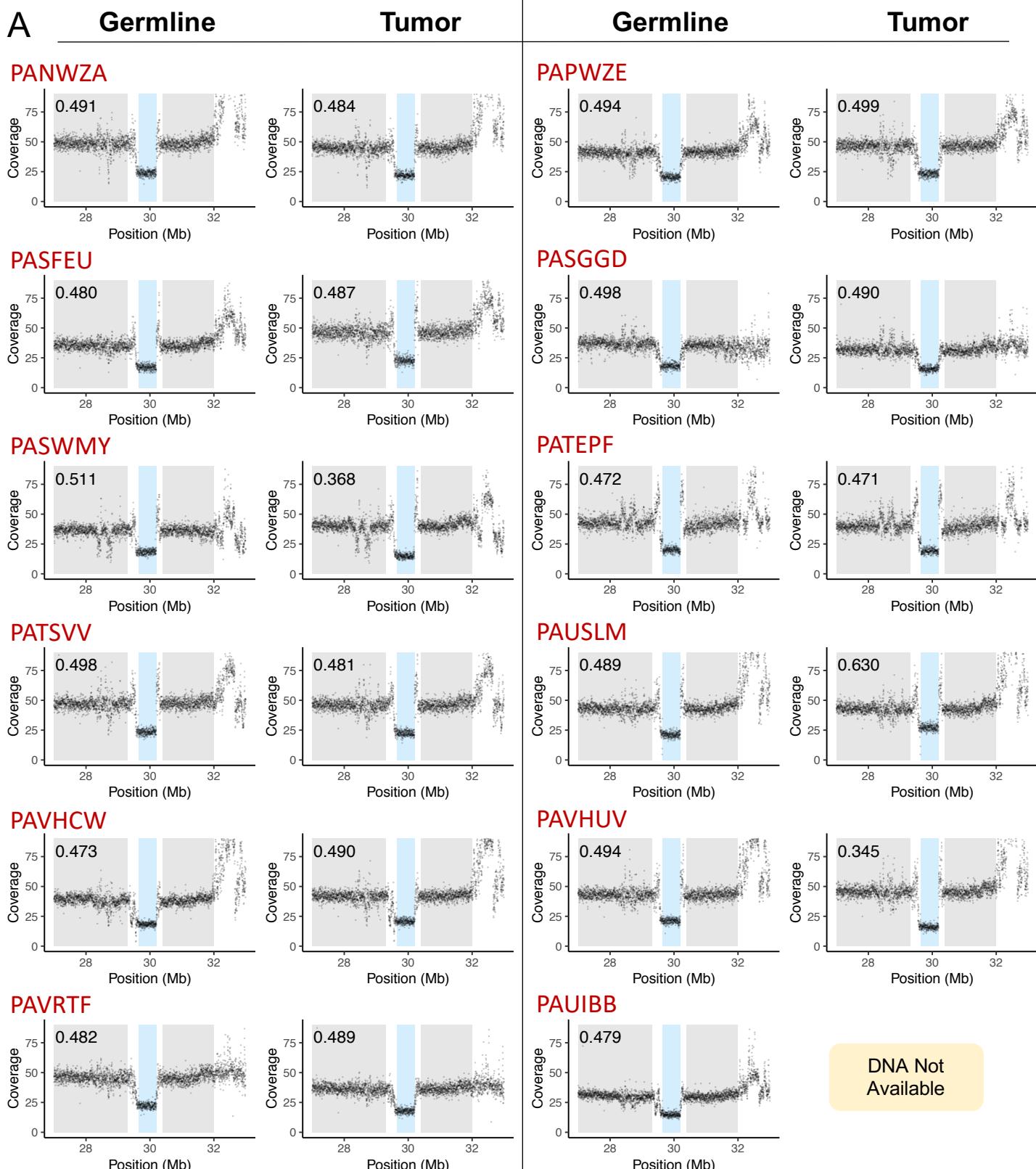
**Figure S2. Inferred ethnicities for individuals with 16p11.2 microdeletion.** Cases and controls with 16p11.2 microdeletion are highlighted relative to discovery (left) and replication (right) cohorts. The frequency of microdeletions within each inferred population is shown (bottom). See Figure S1 for further detail about ethnicity inference.



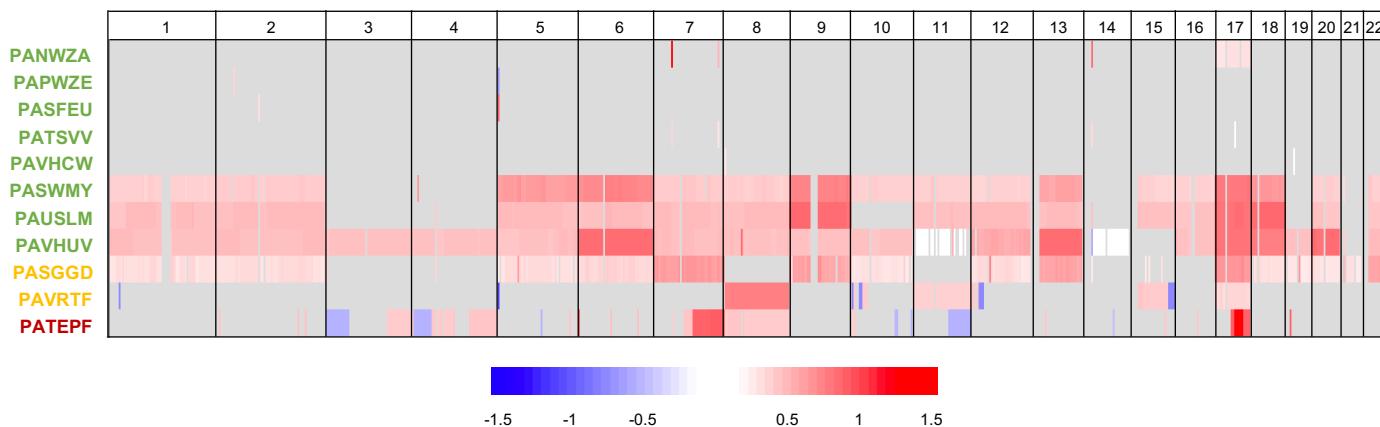
**Figure S3. Megabase-scale germline chromosomal abnormalities in cancer-free controls.** Germline deletions and duplications larger than 5 Mb on autosomal chromosomes are plotted in a summary karyotype showing the entire cohort of 23,505 cancer-free controls (see Figure 1 for neuroblastoma cases). Partial or complete trisomy 21 was observed in 54 controls, so for ease of visualization only 4 are shown. Although controls were filtered to remove severe phenotypes including cancer, neurological disorders, and immunological disorders prior to inclusion in the study, several controls show abnormalities in regions associated with genomic disorders, such as the Prader-Willi critical region on 15q11.2.

**A****B**

**Figure S4. Survival among low-risk neuroblastoma cases with and without 16p11.2 microdeletion.**  
Survival curves showing overall survival (A) and event-free survival (B) in low-risk neuroblastoma cases with 16p11.2 microdeletion (blue, n=14) and without the deletion (orange, n=1,732). Differences between the curves were tested using the log-rank test.



**Figure S5. Read depth across the 16p11.2 region in germline and tumor DNA from 12 neuroblastoma cases with 16p11.2 microdeletion.** A) Raw coverage plotted in 2-kb bins for germline and tumor DNA from cases with 16p11.2 microdeletion. The 16p11.2 deletion critical region (29.65–30.20 Mb) is shaded in blue. Relative coverage within the critical region is displayed in the upper left corner of each plot and was calculated as average coverage across the critical region divided by average coverage across the surrounding region (27.0–29.3 and 30.4–32.0 Mb, shaded in gray). B) Distribution of relative coverage within the 16p11.2 deletion across all whole-genome sequencing samples, including patient germline and tumor DNA (coverage plots shown in Panel A) and parental germline DNA (coverage plots shown in Figure 3). The parental samples form a peak at 1 (no copy number change), and all patient samples except three tumors form a peak at 0.5 (1-copy deletion).



**Figure S6. Whole-genome copy number profiles for tumor DNA from 11 neuroblastoma cases with 16p11.2 microdeletion.** Regions with amplification are shown in red and deletion in blue. Regions with no copy number or allelic change are gray, while regions with copy-neutral loss of heterozygosity are shown in white. Sample names are color-coded by their risk group (green=Low, yellow=Intermediate, red=High). The risk group for PAVRTF was missing in our clinical annotation, but we labeled this patient as intermediate-risk based on INRG stage M, no MYCN amplification, and diagnosis before 12 months of age.

		Discovery				Replication			Both Cohorts	
		HH550v1 <sup>a</sup>	HH550v3	H610-Quad	Total	Omni12v1	Omni24v1	Total		
Number of samples in pre-QC set	Case	1167	1037	1484	<b>3688</b>	1227	1157	<b>2384</b>	<b>6072</b>	
	Control	2324	4993	1912	<b>9229</b>	15013	0	<b>15013</b>	<b>24242</b>	
Number of samples in post-QC set	Case	918	928	1463	<b>3309</b>	1221	1055	<b>2276</b>	<b>5585</b>	
	Control	2098	4888	1869	<b>8855</b>	14650	0	<b>14650</b>	<b>23505</b>	
Number of samples removed	Nexus quality score >0.1	Case	241	47	16	<b>304</b>	5	101	<b>106</b>	<b>410</b>
		Control	223	101	41	<b>365</b>	336	0	<b>336</b>	<b>701</b>
	Wavy or noisy signal	Case	0	1	0	<b>1</b>	1	0	<b>1</b>	<b>2</b>
		Control	3	4	2	<b>9</b>	27	0	<b>27</b>	<b>36</b>
	Tumor contamination	Case	8	61	5	<b>74</b>	0	1	<b>1</b>	<b>75</b>
		Control	0	0	0	<b>0</b>	0	0	<b>0</b>	<b>0</b>

<sup>a</sup> Sample counts are subdivided by the SNP array type: HumanHap550-v1 and v3 (HH550v1 and HH550v3), Human610-Quad (H610-Quad), HumanOmniExpress-12v1 (Omni12v1), and HumanOmniExpress-24v1 (Omni24v1).

**Table S1. Sample sizes before and after quality control (QC) and number of samples removed during each QC step.**

Gene	Location	Top SNP <sup>a</sup>	Cases without 16p11.2 microdeletion			Cases with 16p11.2 microdeletion			P-value (Fisher's exact test on allele counts)	FDR-adjusted P-value <sup>b</sup>	Odds Ratio (95% Confidence Interval)
			C/C	C/T	T/T	C/C	C/T	T/T			
HSD17B12	11p11	rs10742682	0.270 895	0.505 1676	0.225 745	0.231 3	0.462 6	0.308 4	0.561	0.750	1.277 (0.547-3.027)
BARD1	2q35	rs2070096	0.508 1715	0.399 1346	0.093 313	0.800 12	0.200 3	0.000 0	0.025	0.134	0.269 (0.052-0.876)
LMO1	11p15	rs2168101	0.639 1827	0.325 928	0.036 102	0.364 4	0.636 7	0.000 0	0.178	0.406	1.889 (0.650-4.935)
TP53	17p13	rs35850753	0.933 3147	0.066 222	0.001 5	1.000 15	0.000 0	0.000 0	0.625	0.750	0 (0-3.703)
CPZ	4p16	rs3796727	0.434 1441	0.449 1491	0.118 391	0.267 4	0.400 6	0.333 5	0.034	0.134	2.198 (1.003-4.874)
CASC15	6p22	rs4712656	0.266 892	0.511 1714	0.223 746	0.214 3	0.571 8	0.214 3	0.852	0.852	1.091 (0.481-2.473)
MLF1	3q25	rs6441201	0.275 928	0.494 1671	0.231 781	0.200 3	0.400 6	0.400 6	0.203	0.406	1.636 (0.744-3.730)
HACE1/ LIN28B	6q16	rs72990858	0.906 2988	0.091 301	0.003 10	0.867 13	0.133 2	0.000 0	0.656	0.750	1.397 (0.161-5.579)

<sup>a</sup> All SNPs presented in this table correspond to the top SNP reported by McDaniel *et al.*<sup>1</sup> except the *BARD1* top SNP, which is not present in our updated imputation. The second top SNP for *BARD1* reported by McDaniel *et al.* is presented here instead.

<sup>b</sup> P-values were adjusted for a 5% false discovery rate (FDR) using the Benjamini-Hochberg procedure.

**Table S5. Genotype frequencies and counts at known neuroblastoma susceptibility loci among European neuroblastoma cases with and without 16p11.2 microdeletion.**

Case Identifier	Gene	Position	dbSNP rs Number <sup>b</sup>	Variant Effect	Ref. / Alt. Allele	Count of Ref. / Alt. Reads		CADD Score <sup>c</sup>	Freq. in gnomAD (all) <sup>d</sup>	HGMD Entry <sup>e</sup>	Inheritance
						Germline	Tumor				
PANWZA	KIF22	29816237	rs146561986	Missense: D594N	G / A	0 / 25	0 / 19	20.4	0.14%	-	N/A
PANWZA	SEZ6L2	29899021	rs117448844	Missense: R386H	C / T	0 / 20	0 / 14	23	0.16%	CM099077: Autism-associated polymorphism	N/A
PAUIBB	PRRT2	29825022	rs76335820	Missense: P216L	C / T	0 / 11	N/A	25.8	0.79%	CM146179: Probable/possible causal mutation in Landau-Kleffner syndrome	Inherited SNP from heterozygous mother (de novo 16p11.2 microdeletion occurred on paternal allele)
PAUIBB	MVP	29851582	-	Frameshift	G / GTGGA	10 / 15	N/A	24.1	-	-	De novo insertion (likely mosaic)
PAVHCW	MVP	29853091	rs114581451	Missense: R456C	C / T	0 / 10	0 / 19	28.1	0.69%	-	N/A

<sup>a</sup> Rare variants were defined as those with allele frequency <1%; predicted-damaging variants were selected for high- or moderate-impact classification in SnpEff and Combined Annotation Dependent Depletion (CADD) score ≥20

<sup>b</sup> Reference SNP (rs) number from the Short Genetic Variations database (dbSNP) build 146

<sup>c</sup> PHRED-scaled CADD GRCh37-v1.4 score

<sup>d</sup> Allele frequency across all gnomAD v2.0.1 genomes

<sup>e</sup> Human Gene Mutation Database (HGMD) Professional 2016.1

Ref. = Reference; Alt. = Alternate; Freq. = Frequency

**Table S6. Rare, predicted-damaging germline variants<sup>a</sup> identified in 12 neuroblastoma cases with 16p11.2 microdeletion.**

## SUPPLEMENTAL MATERIAL AND METHODS

### Sample Selection and Genotyping

Cases diagnosed with neuroblastoma or ganglioneuroblastoma were recruited through the Children's Oncology Group (COG) biology study (ANBL00B1) without selection for clinical presentation or ethnicity. Germline DNA was isolated from peripheral blood or bone marrow mononuclear cells at time of diagnosis and genotyped at the Children's Hospital of Philadelphia's Center for Applied Genomics (CHOP CAG) on the following Illumina SNP arrays: HumanHap550v1, HumanHap550v3, and Human610-Quad (discovery cohort); HumanOmniExpress-12v1 and HumanOmniExpress-24v1 (replication cohort). Data from control children recruited through the CHOP Health Care Network and genotyped on the same arrays were obtained from the CHOP CAG after screening for cancer and severe neurological or immunological disorders. Only samples with a genotyping call rate >90% were included in the study. In cases where a pair of samples showed a high degree of relatedness (defined as proportion IBD >0.125 calculated using PLINK 1.9<sup>2</sup>) due to cryptic relatedness or duplicate genotyping runs, the sample with the higher call rate was kept.

### SNP Array Data Processing and Quality Control

SNP array intensity files were processed with GenomeStudio 2.0 to calculate log R ratio and B allele frequency values for each SNP using standard SNP clustering with default settings. Genotype calls were generated separately using a custom CHOP CAG pipeline. SNPs from the HumanHap550 arrays were mapped to human genome assembly GRCh37 (hg19) using UCSC liftOver<sup>3</sup> (all genomic coordinates used throughout this manuscript reference hg19). SNPs with minor allele frequency <1%, call rate <90%, or deviation from Hardy-Weinberg equilibrium surpassing p=5x10<sup>-5</sup> were removed using PLINK 1.9. The analysis was then restricted to SNPs present on all chip types within each cohort (402,743 and 506,045 SNPs for discovery and replication, respectively).

### Ethnicity Inference

Case and control subject ancestries were inferred using principal component analysis in PLINK 1.9. SNPs from the discovery and replication cohorts were intersected with HapMap 3 (draft release 2)<sup>4</sup>. This subset was pruned to remove highly correlated SNPs using a window size of 50 variants, step size of 5 variants, and pairwise r<sup>2</sup> threshold of 0.2. The 18,599 remaining SNPs were used to calculate 20 principal components. The first three components were plotted to establish subject ethnicity cutoffs based on their locations relative to HapMap reported ethnicities.

### CNV Detection

Copy number calls were generated from SNP array data using Nexus Copy Number 8.0.<sup>5</sup> Log R ratios were linearly corrected for GC content waves using the correction files provided by BioDiscovery. The SNPRank Segmentation algorithm was used with all default settings except the following: minimum number of probes per segment 10; gain threshold 0.15; loss threshold -0.15; homozygous frequency threshold 0.95. Only CNV calls larger than 500-kb were considered for downstream analysis.

### Sample-level Quality Control

Samples with a Nexus Quality Score (based on variance of log R ratio) greater than 0.1 or visibly noisy or wavy signals were excluded. To prevent confounding of germline copy number analysis by high levels of circulating tumor DNA, samples showing mosaicism for segmental chromosomal alterations that frequently occur in neuroblastoma (loss of 1p, loss of 11q, or gain of 17q) were identified and excluded. Samples with focal *MYCN* amplification were not excluded, as nearly every patient with somatic *MYCN* amplification showed some level of *MYCN* amplification in the germline due to the high copy number in the tumor. However, this region was removed during downstream analysis as noted below.

## CNV Filtering

Stringent CNV filtering was applied prior to all analyses, with the exception of the megabase-scale CNVs shown in Figure 1 (which were manually evaluated to remove artifacts and potential tumor contamination). For all other analyses, steps were taken to reduce confounding by artifacts. CNV calls that were not supported by evidence in the B allele frequency (either loss of heterozygosity or allelic imbalance) were removed. Additionally, sex chromosomes and the extended region of chromosome 2 often co-amplified with MYCN and ALK (2p25.2 to 2p23.1; 4.4–32.1 Mb) were excluded.

## CNV Regional Association Test

The Nexus comparison tool was used to fragment copy number regions into segments with constant CNV frequency, considering only CNVs larger than 500-kb. Deletions and duplications were analyzed separately. Segments that were deleted or duplicated in at least three cases in the discovery cohort (frequency >0.09%) were selected. Among these, contiguous segments were merged to form copy number variable regions (CNVRs). Recurrent CNVRs reaching at least 100-kb in total length were tested for association with neuroblastoma using a two-sided Fisher's exact test, selecting for the most highly associated peak within each CNVR. The significance threshold was adjusted for multiple testing using Bonferroni correction. Regions reaching at least nominal significance were manually evaluated using the Nexus chromosome viewer to detect false positive CNVs and identify additional true positive CNVs that failed CNV filtering. Regions reaching genome-wide significance were then validated in the independent replication cohort. Meta-analysis of the two cohorts was completed using the *metafor* R package<sup>6</sup> with a linear model in R version 3.5.2.<sup>7</sup>

## Clinical Covariate Analysis

16p11.2 microdeletion was tested for association with *MYCN* amplification, Children's Oncology Group (COG) risk stratification, International Neuroblastoma Staging System (INSS) stage, sex, age at diagnosis, and genotypes at known neuroblastoma-associated SNPs using two-sided Fisher's exact tests due to low cell counts. Tests were implemented using the *fisher.test()* function in R version 3.5.2, which applies an extension of Fisher's exact test to datasets with more than two categories. To test association with age at diagnosis, patients were stratified by a cutoff of 18 months.<sup>8</sup> For genotypes, the test was applied to major and minor allele counts, and p-values were corrected by the Benjamini–Hochberg procedure using *p.adjust()* in R version 3.5.2. Survival analysis was completed using the *survival* and *survminer* packages in R version 3.5.2.

## Whole-genome Sequencing (WGS)

WGS data for twelve neuroblastoma patients carrying 16p11.2 microdeletion were obtained from the following sources:

- WGS data for one carrier (case identifier: **PATEPF**) were obtained from TARGET (see <https://ocg.cancer.gov/programs/target/data-matrix> for details). Coverage data, copy number calls, and variant calls for this cohort were generated by Complete Genomics.
- WGS data for four carriers (case identifiers: **PASFEU**, **PASGGD**, **PASWMY**, **PAUIBB**) and their parents were obtained through the Gabriella Miller Kids First (GMKF) program. Blood or buccal DNA was provided for GMKF by the Neuroblastoma Epidemiology in North America (NENA) study.
- WGS data were newly generated for seven additional carriers (case identifiers: **PANWZA**, **PAPWZE**, **PATSVV**, **PAUSLM**, **PAVHCW**, **PAVHUV**, **PAVRTF**). Blood DNA was obtained from the Children's Oncology Group (COG) biobank.

The four carriers from GMKF and seven newly sequenced carriers were sequenced and analyzed using the following methods:

Prior to library preparation, DNA quality was assessed by fluorometric assay for concentration and pulse-field gel electrophoresis for integrity. DNA was then fragmented and sequenced to 30x mean coverage on an Illumina HiSeq 10X, generating 150 bp paired-end reads. FastQC was applied for read quality assessment.

Alignment and germline variant calling were completed using a public CAVATICA workflow (<https://cavatica.sbggenomics.com>; "Whole Genome Sequencing - BWA + GATK 4.0 (with Metrics)," Revision 41) based on the Broad Institute's GATK Best Practices. Briefly, reads were aligned to hg19 with BWA-MEM 0.7.17 and sorted using Samtools. Duplicate reads were removed using Picard. After base quality score recalibration (BQSR), variants were called using HaplotypeCaller. Variant quality score recalibration (VQSR) was then completed using a custom pipeline. VQSR was applied separately to SNPs and indels using several training files available from the GATK Resource Bundle (hapmap3.3 sites, 1KG\_omni2.5, 1KG\_snp/indel\_high.confidence, and dbsnp\_135 files). Passing SNPs and indels were defined as those with a confidence level greater than 99.7 and 99.5, respectively.

Somatic SNVs and indels were called from BAM files using a second public CAVATICA workflow ("VarScan2 Workflow from BAM," Revision 33) with VarScan2 version 2.3.9.<sup>9</sup> Only high confidence SNVs and indels were considered.

Germline and somatic variants were then functionally annotated using SnpEff 4.3. Variants were also annotated using ANNOVAR 2016Feb01 with the following datasets: refGene, knownGene, ensGene, 1000g2015aug\_eas, 1000g2015aug\_eur, 1000g2015aug\_sas, 1000g2015aug\_afr, 1000g2015aug\_all, exac03nontcga, gnomad\_genome, kaviar\_20150923, clinvar\_20170905, avsnp150, cosmic70, dbnsfp33a, dbscsv11, intervar\_20180118, wgRNA, genomicSuperDups, and dgvMerged. Indels longer than 10 base pairs in length were removed. Additionally, variants supported by less than 10 reads or less than 20% of the total reads were removed.

Somatic CNVs were called from BAM files using Nexus Copy Number 10.0<sup>10</sup> with the BAM ngCGH (matched) algorithm and default settings. For hyperploid samples, diploid chromosomes were selected via evaluation of coverage and allele frequency patterns, and log ratios were calculated relative to the median of these chromosomes.

## SUPPLEMENTAL REFERENCES

1. McDaniel, L.D., Conkrite, K.L., Chang, X., Capasso, M., Vaksman, Z., Oldridge, D.A., Zachariou, A., Horn, M., Diamond, M., Huo, C., et al. (2017). Common variants upstream of MLF1 at 3q25 and within CPZ at 4p16 associated with neuroblastoma. *PLoS Genet.* 13, e1006787.
2. Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M., and Lee, J.J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* 4, 7.
3. Hinrichs, A.S., Karolchik, D., Baertsch, R., Barber, G.P., Bejerano, G., Clawson, H., Diekhans, M., Furey, T.S., Harte, R.A., Hsu, F., et al. (2006). The UCSC Genome Browser Database: update 2006. *Nucleic Acids Res.* 34, D590-8.
4. The International HapMap 3 Consortium (2010). Integrating common and rare genetic variation in diverse human populations. *Nature* 467, 52–58.
5. BioDiscovery Inc. (2016). Nexus Copy Number 8.0 Discovery Edition.
6. Viechtbauer, W. (2010). Conducting Meta-Analyses in *R* with the **metafor** Package. *J. Stat. Softw.* 36, 1–48.
7. R Development Core Team (2008). R: A Language and Environment for Statistical Computing.
8. Cohn, S.L., Pearson, A.D.J., London, W.B., Monclair, T., Ambros, P.F., Brodeur, G.M., Faldum, A., Hero, B., Ichihara, T., Machin, D., et al. (2009). The International Neuroblastoma Risk Group (INRG) Classification System: An INRG Task Force Report. *J. Clin. Oncol.* 27, 289–297.
9. Koboldt, D.C., Zhang, Q., Larson, D.E., Shen, D., McLellan, M.D., Lin, L., Miller, C.A., Mardis, E.R., Ding, L., and Wilson, R.K. (2012). VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res.* 22, 568–576.
10. BioDiscovery Inc. (2018). Nexus Copy Number 10.0 Discovery Edition.