

A recurrent 16p12.1 microdeletion supports a two-hit model for severe developmental delay

Santhosh Girirajan^{1,30*}, Jill A Rosenfeld^{2,30}, Gregory M Cooper¹, Francesca Antonacci¹, Priscillia Siswara¹, Andy Itsara¹, Laura Vives¹, Tom Walsh³, Shane E McCarthy⁴, Carl Baker¹, Heather C Mefford¹, Jeffrey M Kidd¹, Sharon R Browning⁵, Brian L Browning⁵, Diane E Dickel¹, Deborah L Levy^{6,7}, Blake C Ballif², Kathryn Platky⁸, Darren M Farber⁹, Gordon C Gowans⁸, Jessica J Wetherbee⁸, Alexander Asamoah⁸, David D Weaver¹⁰, Paul R Mark¹⁰, Jennifer Dickerson¹¹, Bhuwan P Garg¹¹, Sara A Ellingwood¹², Rosemarie Smith¹², Valerie C Banks¹², Wendy Smith¹², Marie T McDonald¹³, Joe J Hoo¹⁴, Beatrice N French¹⁴, Cindy Hudson¹⁵, John P Johnson¹⁵, Jillian R Ozmore¹⁶, John B Moeschler¹⁶, Urvashi Surti¹⁷, Luis F Escobar¹⁸, Dima El-Khechen¹⁸, Jerome L Gorski¹⁹, Jennifer Kussmann¹⁹, Bonnie Salbert²⁰, Yves Lacassie²¹, Alisha Biser²², Donna M McDonald-McGinn²², Elaine H Zackai²², Matthew A Dearnorff²², Tamim H Shaikh²², Eric Haan^{23,24}, Kathryn L Friend²⁵, Marco Fichera²⁶, Corrado Romano²⁶, Jozef Gécz^{24,25}, Lynn E DeLisi^{7,27}, Jonathan Sebat²⁸, Mary-Claire King^{1,3}, Lisa G Shaffer² & Evan E Eichler^{1,29}

We report the identification of a recurrent, 520-kb 16p12.1 microdeletion associated with childhood developmental delay. The microdeletion was detected in 20 of 11,873 cases compared with 2 of 8,540 controls ($P = 0.0009$, $OR = 7.2$) and replicated in a second series of 22 of 9,254 cases compared with 6 of 6,299 controls ($P = 0.028$, $OR = 2.5$). Most deletions were inherited, with carrier parents likely to manifest neuropsychiatric phenotypes compared to non-carrier parents ($P = 0.037$, $OR = 6$). Probands were more likely to carry an additional large copy-number variant when compared to matched controls (10 of 42 cases, $P = 5.7 \times 10^{-5}$, $OR = 6.6$). The clinical features of individuals with two mutations were distinct from and/or more severe than those of individuals carrying only the co-occurring mutation. Our data support a two-hit model in which the 16p12.1 microdeletion both predisposes to neuropsychiatric phenotypes as a single event and exacerbates neurodevelopmental phenotypes in association with other large deletions or duplications. Analysis of other microdeletions with variable expressivity indicates that this two-hit model might be more generally applicable to neuropsychiatric disease.

Most known recurrent genomic disorders result from non-allelic homologous recombination (NAHR) or unequal crossing-over between large and highly identical segmental duplications (>10 kb)¹. Specific human chromosomes (for example, 7, 15, 16, 17 and 22) are enriched for interspersed segmental duplications², and multiple genomic disorders have already been assigned to these 'hot spot' regions of the genome^{3,4}. The short arm of chromosome 16 is particularly enriched for large segmental duplications that have arisen specifically during the evolution of humans and great apes⁵⁻⁸. In the last 3 years, at least three microdeletion/microduplication syndromes have been characterized whose breakpoints map within segmental duplications on chromosome 16. These include a 500-kb microdeletion/duplication of 16p11.2 associated with autism-spectrum disorders and intellectual disability^{9,10}, a large microdeletion encompassing 16p11.2–p12.2 in individuals with a wide spectrum of phenotypic features¹¹ and distal 16p13.11 rearrangements in individuals with autism, intellectual disability and other neurodevelopmental phenotypes¹²⁻¹⁴.

We recently performed a genome-wide meta-analysis comparing the frequency of large deletion/duplication events in individuals with neurocognitive or psychiatric disabilities to that in the control population¹⁵. This meta-analysis pointed towards a potential fourth example of a duplication-mediated microdeletion of ~600 kb on chromosome 16p12.1 that was found in 5 of 6,860 individuals affected by schizophrenia or autism, compared with 0 of 5,674 controls¹⁵. Therefore, we sought to systematically characterize this particular microdeletion in two large cohorts (discovery set and replication set) of children with intellectual disability ($n = 11,873$ and 9,254, respectively) compared with controls ($n = 8,540$ and 6,299, respectively). Our data indicate that the 16p12.1 microdeletion might be a risk factor for neurodevelopmental disease that also acts with other large CNVs to modify neuropsychiatric phenotypes, thereby supporting a 'two-hit' model for the generation of severe cognitive deficits involving this region. Analysis of other microdeletions indicates that this model might help to explain the variability in expressivity of recurrent CNVs associated with neuropsychiatric phenotypes.

*A full list of author affiliations appears at the end of the paper.

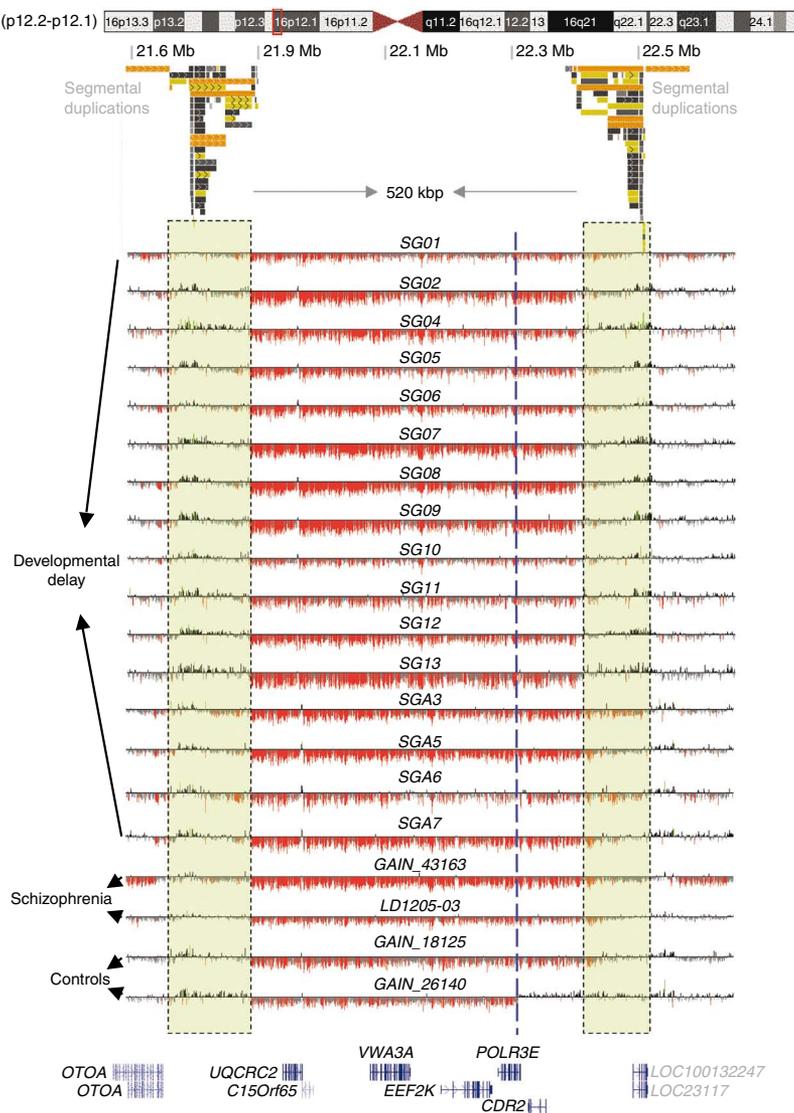


Figure 1 High-resolution array-based CGH characterization of 16p12.1 microdeletion. Shown is validation of 16p12.1 microdeletions in a representative set of cases using high-resolution tiling-path custom array-based CGH. Probes with \log_2 ratios above or below a threshold of 1.5 s.d. from the normalized mean \log_2 ratio are colored green (duplication) or red (deletion), respectively. Dotted lines represent breakpoint regions. Subjects SG01–13 and SGA3–7 have indications of developmental delay or cognitive disability, sample 43163 is from the GAIN schizophrenia study and subject LD1205-03 has schizophrenia and intellectual disability (from family LD1205). Note that samples 26140 and 18125 were analyzed as part of the GAIN control cohort for schizophrenia. It is noteworthy that one control (subject 26140) was retrospectively diagnosed with a major depressive disorder. Six RefSeq genes map within the 16p12.1 microdeletion. Four subjects (SG04, SG07, SG11 (affected with hypoplastic left heart syndrome) and LD1205-03 (diagnosed with schizophrenia)) were sequenced for *CDR2*, *EEF2K* and *UQCRC2*; no recessive mutations were identified.

and 14,839 controls), the pathogenic association of the 16p12.1 microdeletion was highly significant (Fisher's exact test, $P = 1.18 \times 10^{-4}$, OR = 3.7; **Table 1**).

We also examined 3,061 individuals who had been diagnosed with schizophrenia and identified three 16p12.1 microdeletions, all in individuals with sporadic schizophrenia (**Table 1**). However, we did not observe significant enrichment for the 16p12.1 event specifically in the schizophrenia cases compared to 14,839 total controls (Fisher's exact test, $P = 0.29$, OR = 1.8), although this might indicate a lack of statistical power in the schizophrenia panel rather than a true lack of disease association (only 3,061 individuals, in contrast to the 21,127 individuals with developmental delay). It is interesting, in this regard, that in one family where both schizophrenia and mental retardation phenotypes were segregating, the 16p12.1 deletion was observed only among individuals diagnosed with both psychosis and severe intellectual disability (see **Supplementary Note**, family LD1205).

Using high-density and targeted array-based comparative genomic hybridization (CGH) experiments (see Online Methods), we mapped the 16p12.1 microdeletion breakpoints in 37 individuals to two large blocks of segmental duplications (**Fig. 1** and **Supplementary Fig. 1**).

RESULTS

Pathogenic association of 16p12.1 microdeletion

From our initial discovery cohort, we identified 20 affected individuals (cases) with a 520.8-kb 16p12.1 microdeletion (UCSC Human Genome Version Hg18, NCBI Build 36, Chr. 16: 21854025–22374785) among 11,873 individuals with indications of intellectual disability/developmental delay (ID/DD) and congenital malformation (**Fig. 1**; see Online Methods). By contrast, CNV studies on 8,540 controls identified only two individuals, both from the GAIN cohort, with the 16p12.1 microdeletion (**Table 1**). Thus, 16p12.1 microdeletions are significantly enriched in the panel of children with developmental delay studied here (Fisher's exact test, $P = 0.0009$, OR = 7.2). To replicate this association, we evaluated CNV data from an independent set of 9,254 individuals with ID/DD and 6,299 controls (see Online Methods). The microdeletion was identified in 22 of 9,254 cases and 6 of 6,299 controls, confirming a significant enrichment of 16p12.1 microdeletion in the affected individuals (Fisher's exact test, $P = 0.028$, OR = 2.5; **Table 1** and **Supplementary Fig. 1**). For the combined set (21,127 cases

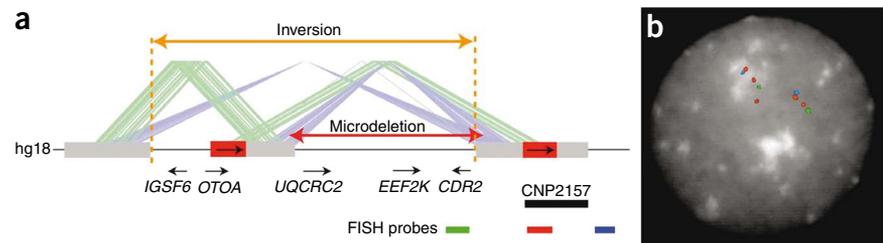
Table 1 Frequency of 16p12.1 microdeletion in cases and controls

	Cases		Controls		Significance	
	del 16p12.1	Total	del 16p12.1	Total	<i>P</i> value	OR
ID/DD cohort						
Discovery set ^a	20	11,873	2	8,540	0.0009	7.2
Replication set ^b	22	9,254	6	6,299	0.028	2.5
Combined ^c	42	21,127	8	14,839	0.000118	3.7
Schizophrenia cohort						
	3	3,061	8	14,839	0.29	1.8

^aThe 16p12.1 microdeletion was originally identified from a meta-analysis that identified 5 cases from 6,860 individuals with autism, schizophrenia and developmental delay compared to 0 observations in a control group of 5,674 ($P = 0.049$). None of these cases were used in the discovery set, although the original control group was expanded in the discovery set. ^bFor the replication set, all controls and cases were independent. ^cCombined set composed of discovery set and replication set.

Figure 2 Genomic structure of 16p12.1 region.

(a) A schematic of the 16p12.1 region shows the location of the microdeletion flanked by directly oriented 68-kb segmental duplication blocks (red boxes). The segmental duplication blocks (red and gray boxes) are connected by green and blue lines to indicate direct or inverted orientation, respectively. Also shown are representative genes in the region with the transcriptional direction. CNP indicates the copy-number polymorphism annotated by SNP genotyping¹⁶ for this region (CNP2157). (b) FISH analysis was performed on lymphoblast cell line from GM18956 using fosmid probes mapping to the 68-kb duplicon (WIBR2-2031K01, shown in red) and the flanking unique regions (WIBR2-3632J22 in green and WIBR2-1829F15 in blue). Results show that the 68-kb duplicon is polymorphic (that is, it has a variable number of copies) and that the orientation of the region is inverted in this HapMap sample compared to the human genome reference assembly. High copy numbers of the segmental duplications have complicated mapping of the inversion breakpoint for this region.



Within these blocks, we identified a 68-kb duplicon in direct orientation (with 99.5% identity) in relation to its paralog in the distal breakpoint region (**Supplementary Table 1**). As misalignment of directly oriented duplicons during meiosis predisposes to NAHR events resulting in microdeletions¹, the 68-kb duplicon probably mediates the observed recurrent 16p12.1 rearrangements. Numerous studies of the region using FISH (**Fig. 2**), SNP microarrays¹⁶ and sequencing-based approaches¹⁷ reveal that the 68-kb duplicon varies in copy number and that the entire region can be inverted in some individuals, as observed for the 17q21.31 microdeletion¹⁸. Preliminary analyses suggest that particular structural configurations might be more predisposed to 16p12.1 rearrangement (data not shown). Owing to the complexity at this locus and the high copy-number variation of the duplications, we could not refine the boundary of the breakpoints below 100 kb by array-based CGH. Notably, the microdeletion in one control sample (GAIN_26140) was 102.9 kb shorter in length than those found in affected subjects and did not span the *CDR2* gene (encoding cerebellar degeneration related 2) (**Fig. 1**).

Phenotypic evaluation and parental analysis

Evaluation of available medical records from the ID/DD cohort showed that multiple phenotypic features are associated with the 16p12.1 microdeletion (**Table 2** and **Supplementary Tables 2** and **3**). Although not all clinical features were completely recognizable in very young subjects, most subjects showed developmental delay and learning disability. All 15 individuals who were older than 12 months had speech delay. Craniofacial and skeletal abnormalities were observed in 22 of 23 cases (**Fig. 3**). Growth retardation was documented in 9 of 22 cases and 7 of 20 individuals also had microcephaly (**Table 2**).

(Note that the denominator of these data varies because not all patients could be ascertained for all clinical features.) Furthermore, congenital cardiac disease was observed in 7 of 21 cases, of which four were specifically diagnosed with a hypoplastic left heart syndrome (**Table 2** and **Supplementary Table 3**). Seizure disorders manifesting in various forms, including West syndrome, febrile seizures or seizure-like episodes, were observed in 8 of 22 cases, and hypotonia was present in 10 of 21 cases (**Table 2**). Psychiatric and behavioral abnormalities were also documented in 9 of 16 affected children. Non-typical facial gestalt and variable clinical presentations indicate that this microdeletion is nonsyndromic.

Six of twenty individuals with a 16p12.1 microdeletion from the discovery set had an additional chromosomal abnormality or large (>500 kb) CNV (**Fig. 4**, **Table 3** and **Supplementary Table 4**). The frequency (30%) of such double-hit CNVs was 7.5-fold greater in the cases with 16p12.1 microdeletion (Fisher's exact test, $P = 0.0005$, OR = 9.7) than in controls conditioned for a large CNV (deletion or duplication) first hit (9 of 217 or 4.1%; **Table 3**). In the replication set, the double-hit frequency was also enriched in affected individuals (4 of 22, 18.2%) compared to controls (12 of 254, 4.7%; Fisher's exact test, $P = 0.029$, OR = 4.5; **Supplementary Table 2** and **Supplementary Fig. 2**). Overall, in the combined set, there was a significant (Fisher's exact test, $P = 5.7 \times 10^{-5}$, OR = 6.6) enrichment of double hits among 16p12.1 deletion carriers (10 of 42, 24%) compared to controls (21 of 471, 4.4%; **Table 3**). Significant enrichment for double hits in cases with 16p12.1 microdeletion was also observed when the controls were conditioned to carry only a large deletion event as the first hit (**Supplementary Table 5**).

In the cases where the second-hit CNV is associated with a described syndrome (DECIPHER¹⁹ and ECARUCA²⁰ databases or published

Table 2 Frequency of phenotypic features in individuals with 16p12.1 deletions

	01	02	03	04	05	06	07	08	09	10	11	12	13	A1	A2	A3	A5	A6	A7	A8	A9	A28	25514	Frequency	%
Developmental delay	+	+	+	N	+	+	+	+	+	+	N	+	+	+	+	+	+	N	N	N	+	+	+	18/18	100%
Speech delay ^a	+	+	+	N	+	+	+	+	+	+	N	N	+	N	+	+	+	N	N	N	+	N	+	15/15	100%
Craniofacial, skeletal features	+	+	+	+	+	+	+	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	22/23	96%
Growth retardation	-	+	-	-	-	+	-	+	+	+	+	N	-	+	-	-	-	-	-	-	+	+	-	9/22	41%
Microcephaly	+	+	-	-	-	+	+	+	+	-	-	N	-	N	-	-	-	-	-	-	N	+	-	7/20	35%
Congenital cardiac defect	-	-	-	+	-	-	+	-	+	-	+	N	-	N	-	-	-	+	+	+	-	-	-	7/21	33%
Hypoplastic heart	-	-	-	+	-	-	-	-	-	-	+	N	-	N	-	-	-	-	+	+	-	-	-	4/21	19%
Seizures	+	+	-	+	+	-	+	-	-	-	-	+	-	N	-	+	-	+	+	-	-	-	-	8/22	36%
Psychiatric/behavioral features	-	+	-	N	-	+	-	+	+	-	N	N	+	N	+	-	-	N	N	N	+	+	+	9/16	56%
Hearing loss	-	-	-	N	N	-	-	-	-	N	-	N	-	N	-	-	+	+	-	N	-	+	-	3/17	18%
Hypotonia	+	+	+	-	+	-	-	+	+	-	-	N	-	N	+	+	-	-	-	-	+	-	+	10/21	48%
Sacral dimple or tethered cord	-	-	-	-	-	-	-	-	+	+	+	N	-	N	-	-	-	-	+	-	-	-	-	4/21	19%

Gray shaded columns represent individuals who carry 'two hits'. +, phenotype present; -, phenotype absent; N, phenotype not assessed.

^aSpeech delay was not evaluated in very young subjects. ^bPatent ductus arteriosus and patent foramen ovale resolved on a follow-up examination. ^cSlowing on EEG. ^dMyoclonus.

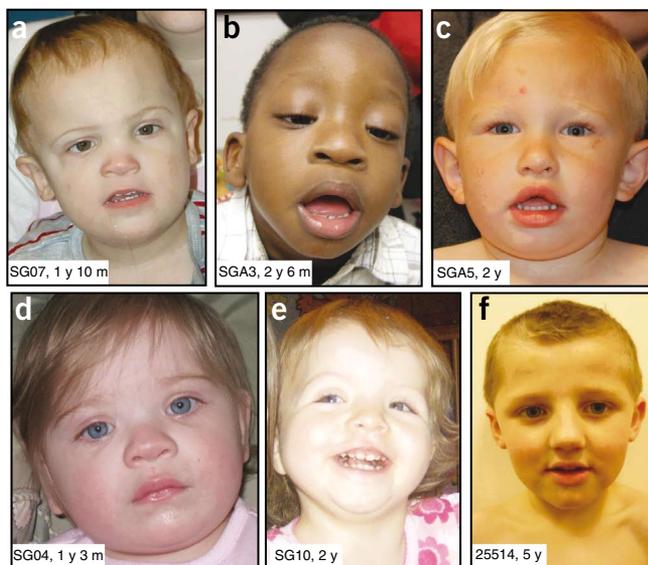


Figure 3 Representative photographs of individuals with 16p12.1 microdeletion. (a–f) Facial features of patient SG07 at 22 months (a), patient SGA3 at 2.5 years (b), patient SGA5 at 2 years (c), patient SG04 at 15 months (d), patient SG10 at 2 years (e) and patient 25514 at 5 years (f). Specific consents were obtained to publish these photographs.

literature), two-hit 16p12.1 carriers manifested more severe or distinct phenotypes than the typical features of the syndrome (**Supplementary Note**). For example, patient SG03 carries a 16p12.1 microdeletion in addition to a 22q13 terminal deletion. The single del22q13ter event has been reported in Phelan-McDermid syndrome²¹ and is associated with autism-spectrum disorders²², developmental delay, hypotonia, normal to accelerated growth and dysmorphic features including macrocephaly, dolichocephaly and micrognathia. However, SG03 manifests with learning disability, microcephaly, exotropia and periventricular abnormalities, with no autistic features yet apparent (age 2 years 3 months). SG10 carries a 35-Mb deletion on chromosome 5q15q23.2 along with the 16p12.1 microdeletion. Previously, Lindgren and colleagues reported on two individuals carrying a del5q15q23.2 with craniofacial features and benign adenomatous polyps in the intestine²³. SG10 has developmental delay, significant growth retardation, overt craniofacial features (ptosis, hypertelorism, bifid uvula, exotropia, ectopic pupils, wide nasal bridge and smooth philtrum), hypoactivity and seizures—a more severe phenotype than that reported from a single second hit. Patient SGA2 carries a 1.3-Mb duplication on chromosome 2p13p12 in addition to the 16p12.1 microdeletion and has overt clinical features, including impaired intellectual function, behavior problems, growth anomalies, craniofacial features and café-au-lait spots. Although dup2p13p12 has been reported to be associated with primary cutaneous lymphoma²⁴, there have been no reports of developmental defect or craniofacial anomalies. Subject SGA6 carries an 848-kb duplication on chromosome 14q32.1 along with the 16p12.1 microdeletion. Additionally, he has had a mutation identified in the *BRAF* gene (F468S) that is consistent with a diagnosis of cardiofaciocutaneous syndrome (CFCS). The subject presents with a diverse range of severe clinical features including craniofacial anomalies, complete agenesis of corpus callosum, renal and cardiac defects and Hirschsprung disease (**Supplementary Table 3**). These features are more severe than has been described for CFCS²⁵ or a *de novo* dup14q32.1 case reported in association with schizophrenia²⁶.

To test whether the subjects inherited the 16p12.1 microdeletion from their parents, we obtained DNA from 23 sets of parents. The 16p12.1 microdeletion was inherited in 22 of 23 cases (17 maternal, 5 paternal), with one case confirmed as being *de novo* (**Fig. 4** and **Supplementary Figs. 3** and **4**). Of the seven double-hit cases where inheritance could be assessed, 6 of 7 large CNV second hits were inherited and one large CNV arose *de novo*. In most cases (5 of 7), either the large CNV second hit was inherited from the parent who did not carry a 16p12.1 microdeletion or it arose *de novo*. In one case (SG12), a heterozygous dup22q11.21 was transmitted from both parents to the homozygote proband (**Fig. 4**). SGA30 inherited both events from the mother (**Supplementary Table 6**).

Parental phenotypic information was obtained during the initial visit before CNV testing from 11 of 13 inherited cases. In two cases, clinical inquiries were made only after the child was found to carry the microdeletion (**Supplementary Table 7**). Carrier parents with the 16p12.1 microdeletion were more likely ($P = 0.037$, OR = 6, Fisher's exact test) to present with learning disability, depression or bipolar disorders, or seizures than the non-carrier parents (**Fig. 4** and **Supplementary Table 8**). We found only three cases (SG11, SGA9 and 25514) in whom the microdeletion was inherited from an apparently normal parent (**Supplementary Table 9**). Compared to the carrier parents, the index cases manifested with more severe, clinically recognizable manifestations, including severe speech and motor disability, recognizable facial dysmorphology and systemic organ abnormalities (**Table 2**, **Figs. 3** and **4** and **Supplementary Note**). Finally, we assessed the SNP genotypes for chromosome 16p in 17 probands and 4 available parents to investigate whether there is one specific sequence haplotype that predisposes to this microdeletion. We found that the microdeletion occurred on different haplotype backgrounds, consistent with a recurrent nature for this rearrangement as opposed to inheritance from a common ancestor (**Supplementary Note**).

DISCUSSION

Previously, we undertook a large meta-analysis study of CNVs in individuals diagnosed with intellectual disability, autism and schizophrenia, identifying the 16p12.1 region as a potential pathogenic locus (five cases and no controls)¹⁵. However, this study lacked the power to definitively identify a disease association or to delineate the phenotypic consequences of the microdeletion. Here we used the CNV data from one of the largest collections of individuals with intellectual disability and developmental delay to reassess this region. We identified 42 index cases with developmental delay, craniofacial dysmorphology and congenital heart defects and three sporadic and one familial case of overt schizophrenia carrying a similar-sized microdeletion of 16p12.1; we found only eight control individuals carrying a 16p12.1 deletion. Although all cases appear to share the segmental duplication-mediated 520-kb minimal deletion region, it is noteworthy that one deletion-carrying control was retrospectively diagnosed with a major depressive disorder and also had an atypical, smaller deletion (417 kb). Removal of this control sample from our analysis strengthens the association between 16p12.1 microdeletion and disease (Fisher's exact test, $P = 4.4 \times 10^{-5}$, OR = 4.2). Comparison to other genomic disorders in our cohort indicates that the incidence of 16p12.1 microdeletion is $\sim 1/15,000$ live births, similar to that of Smith-Magenis syndrome (**Table 4**). However, the subtle clinical features and variable phenotypes associated with the single 16p12.1 event probably lead to underascertainment or misdiagnoses.

Most of our pediatric cases had indications of developmental delay and learning disability and congenital abnormalities. However, the variable phenotypes associated with the 16p12.1 microdeletion

include congenital heart defects, seizures and severe growth abnormalities (**Supplementary Note**). We also identified five adults with a diagnosis of schizophrenia and found that 23% of the probands inherited the microdeletion from a carrier parent with manifestations of psychiatric disease. Carrier parents were significantly more likely ($P = 0.037$) to manifest other neurological or neuropsychiatric features (learning disability, depression, bipolar disorders and seizures)

than non-carriers. The fact that these parents were recruited as a result of their child's referral constitutes an ascertainment bias; however, any such bias should apply to both parents equally, and yet only 5 of 14 non-carrier parents had any clinical symptoms (**Supplementary Table 9**). A detailed clinical examination is therefore warranted in parents of individuals with disorders such as 16p12.1 microdeletion to understand the relationship between the various neurodevelopmental

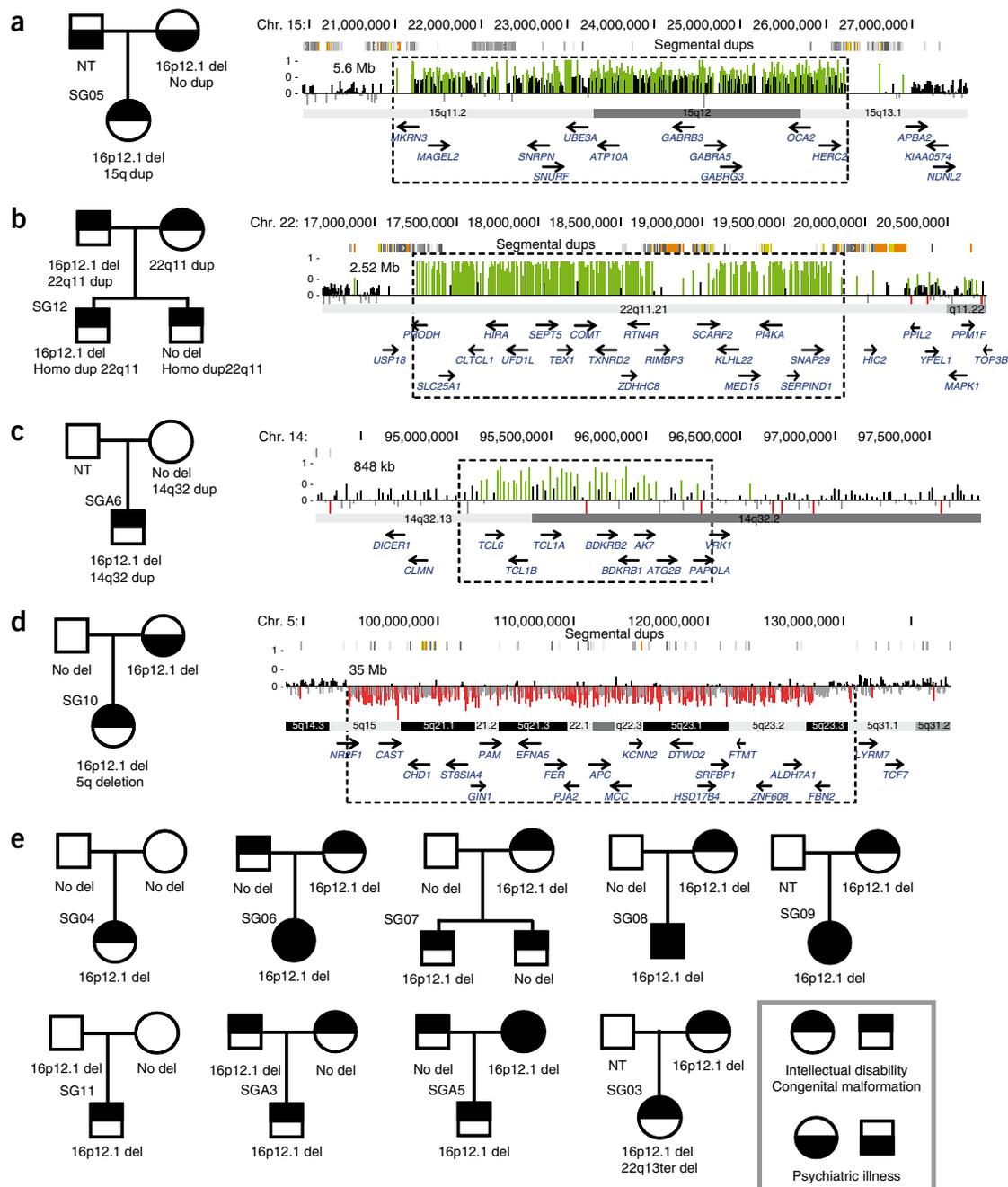


Figure 4 Family pedigrees of probands with 16p12.1 microdeletions. (**a–d**) Large CNVs outside the 16p12.1 region in a representative set of individuals with 16p12.1 microdeletions. The CNV regions are indicated by dotted lines, and the cytogenetic extent and size are labeled. We used a 135K NimbleGen array to identify these CNVs (with average probe density of 2.5 kb in regions flanked by segmental duplications and an average probe density of 35 kb in the genomic backbone). CNV calls were made using a Hidden Markov Model CNV-calling algorithm described previously¹⁵. (**e**) Pedigrees of individuals with 16p12.1 microdeletions and known available parental information. Circles indicate females; squares indicate males. The intellectual disability and congenital malformation category also includes congenital heart defects and seizures. Psychiatric illness includes depression or bipolar disorder, attention deficit hyperactive disorder and abnormal behavior. Note that there is an excess of transmitting parents with the microdeletion who also manifested a neuropsychiatric phenotype. NT, not tested.

Table 3 Enrichment for 'second-hit' CNVs among 16p12.1 microdeletion carriers

	Two large CNVs	Total	Percent	Significance	OR
Discovery cases	6	20	30%	$P = 0.0005$	9.7
Discovery controls ^a	9	217	4.1%		
All discovery controls	9	2,493	0.4%	$P = 0.029$	4.5
Replication cases	4	22	18.2%		
Replication controls ^a	12	254	4.7%		
All replication controls	12	2,792	0.42%	$P = 0.000057$	6.6
Combined cases	10	42	23.8%		
Combined controls ^a	21	471	4.4%		
All controls	12	5,285	0.39%		

^aConditioned on the presence of one large CNV (>500 kb) and finding of a second hit.

and cognitive phenotypes. In probands carrying the microdeletion, we observed a six-fold excess of double CNV hits (10 of 42, 24%) as compared to normal controls when conditioned for the presence of at least one large CNV (21 of 471, 4%) and a more than 60-fold excess as compared to the general population. Compared to the classical phenotypes associated with the known pathogenic locus, each of these children showed additional or more severe phenotypes (**Supplementary Note**). Collectively, these data indicate that phenotypes associated with the 16p12.1 microdeletion show variable expressivity that depends on the genetic background.

We conclude that the deletion of 16p12.1 is a significant, independent risk factor for intellectual disability and developmental delay that also acts with other factors to modify neurological phenotypes. We propose a 'two-hit' model, wherein a secondary insult is necessary during development to result in a more severe clinical manifestation of pediatric disease. The second hit could potentially be another CNV, a disruptive single-base-pair mutation in a phenotypically related gene or an environmental event that influences the phenotype. It has been noted that genetic factors that are originally identified to associate with a specific neurological disease are often subsequently identified in individuals with other conditions (for example, del15q13.3 is associated with epilepsy, schizophrenia and intellectual disability)^{27–29}.

The prevalence of schizophrenia in individuals with a learning disability is reported to be three times that of the general population, and several studies indicate that the comorbidity of schizophrenia in learning-disabled individuals is mainly due to a greater tendency for schizophrenic individuals to develop cognitive delay^{30–32}. The observation, in family LD1205, that cognitive impairment is seen only

among siblings with the 16p12.1 deletion is also consistent with the hypothesis that the 16p12.1 deletion exacerbates the phenotypic consequences of other heritable neurological disease risk factors within this multiplex family (**Supplementary Note**). Notably, about 70% of individuals with autism also present with a learning disability³³. Our 'two-hit' model might also help to explain the significant comorbidity that exists among cognitive impairment, schizophrenia and autism³⁴ and the underlying phenotypic variability reported for several recurrent microdeletions^{29,35}.

To test this, we investigated the occurrence of two hits more broadly among other genomic disorders. We first reanalyzed the recently reported 1q21.1 microdeletion for the presence of a second large CNV³⁶. We found a 40-fold enrichment for two hits (4 of 25, 16%) among cases with 1q21.1 microdeletion when compared to controls. The phenotypes of these four cases were variable, ranging from severe neurological deficit and craniofacial abnormalities to severe schizophrenia without cognitive impairment (**Supplementary Note**). On the basis of this finding, we expanded our analysis to consider both syndromic and non-syndromic genomic disorders for the frequency of double hits. Our analysis for two hits in nine genomic disorders shows that 16p12.1 microdeletion ranks as the top recurrent CNV that is enriched for the occurrence of two hits (**Table 4**). Notably, the proportion of cases with a second hit is generally much higher (9–24% of the cases) for recurrent microdeletions or microduplications where variable expressivity has been reported, including del15q13.3, del16p11.2, dup22q11.2 and del16p12.1. We find an inverse correlation between the proportion of cases that are *de novo* and the prevalence of the second hit. For example, only one *de novo* case (~4%) of the 16p12.1 microdeletion has been reported, and this microdeletion shows the greatest fraction of double hits. By contrast, we observed either few or no double hits among canonical syndromes such as Williams and Smith-Magenis syndromes (**Table 4**), for which almost all cases arise *de novo*. We note, however, that we did not find a significant enrichment for two hits among cases with 22q11.2 deletion—similar to that observed by Bassett and colleagues³⁷. In general, these findings provide additional support for the two-hit hypothesis, with elevated double-hit rates among pathogenic CNVs with clearly variable penetrance and expressivity. We propose that variable phenotypes of microdeletions such as those at 16p12.1 and 1q21.1 are subject to substantial modification (or that these copy-number variants are themselves modifiers).

METHODS

Methods and any associated references are available in the online version of the paper at <http://www.nature.com/naturegenetics/>.

Table 4 Percentage of second hits and *de novo* rates of microdeletions

Disorder	Cases	Total cases	Incidence (%)	<i>De novo</i>	<i>De novo</i> (%)	Second hits	Second hit (%)	<i>P</i> -value
Smith-Magenis syndrome	25	20,647	0.12	3/3	100%	0	0%	–
17q21.31 deletion	29	20,647	0.14	7/7	100%	0	0%	–
Williams syndrome	60	20,647	0.29	5/5	100%	3	5%	0.52
DiGeorge syndrome	113	20,647	0.54	13/17	76.4%	9	8%	0.10
15q13.3 deletion	66	20,647	0.32	6/25	24%	6	9.1%	0.1
16p11.2 deletion	91	20,647	0.44	26/35	74.3%	9	9.9%	0.038
1q21.1 deletion	98	25,866	0.37	15/45	33%	11	11.2%	0.012
22q11.2 duplication	61	20,647	0.30	2/21	9.5%	9	14.8%	0.003
16p12.1 deletion	42	21,127	0.20	1/23	4.3%	10	23.8%	5.7×10^{-5}
Controls ^a	471	5,285	–	–	–	21	4.4%	–
Controls (unconditioned) ^b	–	5,285	–	–	–	21	0.39%	–

^aFor comparison, controls were conditioned to have at least one large CNV (>500 kb) and then the number of second hits in these cases were counted. ^bControls were not conditioned for first hit. This gives an estimate of two hits in the general population compared to affected individuals.

Accession code. dbGaP (genotype data): GAIN study of schizophrenia (phs000021.v2.p1).

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

We thank the subjects and their families for participating in this study. We thank M. Whyte for providing cardiological evaluation (for case SG07). We thank A. Singleton, L. Ferrucci and R. Krauss for sharing the control CNV data generated with support from the Intramural Research Program of the National Institute on Aging, National Heart, Lung, and Blood Institute (PARC project), and the National Institutes of Health, Department of Health and Human Services. For the schizophrenia study, genotyping of samples was also provided through the genetic association information network (GAIN). Samples and associated phenotype data

for the genome-wide association of schizophrenia study were provided by the Molecular Genetics of Schizophrenia Collaboration (principal investigator: P.V. Gejman, Evanston Northwestern Healthcare (ENH) and Northwestern University, Evanston, Illinois, USA). We also thank J. Smith for SNP genotyping; T. Louie for bioinformatics support; C. Campbell, L. Perez-Jurado, G. Kirov, M.K. Rudd and C. Lese-Martin for useful discussions; and T. Brown, P. Sudmant and J. Kitzman for critical review of the manuscript. This work was supported by NIH grant HD065285 (E.E.E.), a grant from the Simons Foundation (SFARI 137578 to E.E.E.) and a NARSAD Award (M.-C.K.). G.M.C. is supported by a Merck, Jane Coffin Childs Memorial Fund post-doctoral fellowship. E.E.E. is an investigator of the Howard Hughes Medical Institute.

AUTHOR CONTRIBUTIONS

This study was designed by S.G., L.G.S. and E.E.E. J.A.R., B.C.B. and L.G.S. supervised array-CGH experiments at Signature Genomics. J.A.R. coordinated clinical data collection. T.W., S.E.M., D.E.D., D.L.L., J.S., L.E.D. and M.-C.K. contributed to schizophrenia data collection and analysis. S.G., L.V. and C.B. performed high-density array-CGH experiments. G.M.C. and A.I. analyzed control CNV data. S.G., F.A. and J.M.K. performed genome structure analysis. F.A. performed FISH experiments. S.G. and P.S. sequenced and analyzed candidate genes. S.R.B. and B.L.B. performed haplotype analysis. K.P., D.M.F., G.C.G., J.J.W., A.A., D.D.W., P.R.M., J.D., B.P.G., S.A.E., R.S., V.C.B., W.S., M.T.M., J.J.H., B.N.F., C.H., J.P.J., J.R.O., J.B.M., U.S., L.E.E., D.E.-K., J.L.G., J.K., B.S., Y.L., A.B., D.M.M.-M., E.H.Z., M.A.D., T.H.S., E.H., K.L.F., M.F., C.R. and J.G. provided clinical information. H.C.M. provided 1q21.1 data. S.G., G.M.C., M.-C.K. and E.E.E. contributed to data interpretation. S.G. and E.E.E. wrote the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare competing financial interests: details accompany the full-text HTML version of the paper at <http://www.nature.com/naturegenetics/>.

Published online at <http://www.nature.com/naturegenetics/>.

Reprints and permissions information is available online at <http://npg.nature.com/reprintsandpermissions/>.

- Lupski, J.R. Genomic disorders: structural features of the genome can lead to DNA rearrangements and human disease traits. *Trends Genet.* **14**, 417–422 (1998).
- Marques-Bonet, T., Girirajan, S. & Eichler, E.E. The origins and impact of primate segmental duplications. *Trends Genet.* **25**, 443–454 (2009).
- Mefford, H.C. & Eichler, E.E. Duplication hotspots, rare genomic disorders, and common disease. *Curr. Opin. Genet. Dev.* **19**, 196–204 (2009).
- Zhang, F., Gu, W., Hurler, M.E. & Lupski, J.R. Copy number variation in human health, disease, and evolution. *Annu. Rev. Genomics Hum. Genet.* **10**, 451–481 (2009).
- Eichler, E.E. *et al.* Divergent origins and concerted expansion of two segmental duplications on chromosome 16. *J. Hered.* **92**, 462–468 (2001).
- Johnson, M.E. *et al.* Recurrent duplication-driven transposition of DNA during hominoid evolution. *Proc. Natl. Acad. Sci. USA* **103**, 17626–17631 (2006).
- Johnson, M.E. *et al.* Positive selection of a gene family during the emergence of humans and African apes. *Nature* **413**, 514–519 (2001).
- Loftus, B.J. *et al.* Genome duplications and other features in 12 Mb of DNA sequence from human chromosome 16p and 16q. *Genomics* **60**, 295–308 (1999).
- Kumar, R.A. *et al.* Recurrent 16p11.2 microdeletions in autism. *Hum. Mol. Genet.* **17**, 628–638 (2008).
- Weiss, L.A. *et al.* Association between microdeletion and microduplication at 16p11.2 and autism. *N. Engl. J. Med.* **358**, 667–675 (2008).
- Ballif, B.C. *et al.* Discovery of a previously unrecognized microdeletion syndrome of 16p11.2-p12.2. *Nat. Genet.* **39**, 1071–1073 (2007).
- Hannes, F.D. *et al.* Recurrent reciprocal deletions and duplications of 16p13.11: the deletion is a risk factor for MR/MCA while the duplication may be a rare benign variant. *J. Med. Genet.* **46**, 223–232 (2009).
- Mefford, H.C. *et al.* A method for rapid, targeted CNV genotyping identifies rare variants associated with neurocognitive disease. *Genome Res.* **19**, 1579–1585 (2009).
- Ullmann, R. *et al.* Array CGH identifies reciprocal 16p13.1 duplications and deletions that predispose to autism and/or mental retardation. *Hum. Mutat.* **28**, 674–682 (2007).
- Itsara, A. *et al.* Population analysis of large copy number variants and hotspots of human genetic disease. *Am. J. Hum. Genet.* **84**, 148–161 (2009).
- McCarroll, S.A. *et al.* Integrated detection and population-genetic analysis of SNPs and copy number variation. *Nat. Genet.* **40**, 1166–1174 (2008).
- Kidd, J.M. *et al.* Mapping and sequencing of structural variation from eight human genomes. *Nature* **453**, 56–64 (2008).
- Zody, M.C. *et al.* Evolutionary toggling of the *MAPT* 17q21.31 inversion region. *Nat. Genet.* **40**, 1076–1083 (2008).
- Firth, H.V. *et al.* DECIPHER: Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources. *Am. J. Hum. Genet.* **84**, 524–533 (2009).
- Feenstra, I. *et al.* European Cytogeneticists Association Register of Unbalanced Chromosome Aberrations (ECARUCA); an online database for rare chromosome abnormalities. *Eur. J. Med. Genet.* **49**, 279–291 (2006).
- Phelan, M.C. *et al.* 22q13 deletion syndrome. *Am. J. Med. Genet.* **101**, 91–99 (2001).
- Durand, C.M. *et al.* Mutations in the gene encoding the synaptic scaffolding protein SHANK3 are associated with autism spectrum disorders. *Nat. Genet.* **39**, 25–27 (2007).
- Lindgren, V., Bryke, C.R., Ozcelik, T., Yang-Feng, T.L. & Francke, U. Phenotypic, cytogenetic, and molecular studies of three patients with constitutional deletions of chromosome 5 in the region of the gene for familial adenomatous polyposis. *Am. J. Hum. Genet.* **50**, 988–997 (1992).
- Mao, X. *et al.* Genetic alterations in primary cutaneous CD30+ anaplastic large cell lymphoma. *Genes Chromosom. Cancer* **37**, 176–185 (2003).
- Roberts, A. *et al.* The cardiofaciocutaneous syndrome. *J. Med. Genet.* **43**, 833–842 (2006).
- Xu, B. *et al.* Strong association of *de novo* copy number mutations with sporadic schizophrenia. *Nat. Genet.* **40**, 880–885 (2008).
- Helbig, I. *et al.* 15q13.3 microdeletions increase risk of idiopathic generalized epilepsy. *Nat. Genet.* **41**, 160–162 (2009).
- Sharp, A.J. *et al.* A recurrent 15q13.3 microdeletion syndrome associated with mental retardation and seizures. *Nat. Genet.* **40**, 322–328 (2008).
- Stefansson, H. *et al.* Large recurrent microdeletions associated with schizophrenia. *Nature* **455**, 232–236 (2008).
- Turner, T.H. Schizophrenia and mental handicap: an historical review, with implications for further research. *Psychol. Med.* **19**, 301–314 (1989).
- Moorhead, T.W. *et al.* Voxel-based morphometry of comorbid schizophrenia and learning disability: analyses in normalized and native spaces using parametric and nonparametric statistical methods. *Neuroimage* **22**, 188–202 (2004).
- Sanderson, T.L., Best, J.J., Doody, G.A., Owens, D.G. & Johnstone, E.C. Neuroanatomy of comorbid schizophrenia and learning disability: a controlled study. *Lancet* **354**, 1867–1871 (1999).
- Fombonne, E. Epidemiological trends in rates of autism. *Mol. Psychiatry* **7** Suppl 2, S4–S6 (2002).
- Woodberry, K.A., Giuliano, A.J. & Seidman, L.J. Premorbid IQ in schizophrenia: a meta-analytic review. *Am. J. Psychiatry* **165**, 579–587 (2008).
- International Schizophrenia Consortium. Rare chromosomal deletions and duplications increase risk of schizophrenia. *Nature* **455**, 237–241 (2008).
- Mefford, H.C. *et al.* Recurrent rearrangements of chromosome 1q21.1 and variable pediatric phenotypes. *N. Engl. J. Med.* **359**, 1685–1699 (2008).
- Bassett, A.S., Marshall, C.R., Lionel, A.C., Chow, E.W. & Scherer, S.W. Copy number variations and risk for schizophrenia in 22q11.2 deletion syndrome. *Hum. Mol. Genet.* **17**, 4045–4053 (2008).

¹Department of Genome Sciences, University of Washington School of Medicine, Seattle, Washington, USA. ²Signature Genomic Laboratories, Spokane, Washington, USA. ³Department of Medicine, University of Washington School of Medicine, Seattle, Washington, USA. ⁴Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, USA. ⁵Department of Statistics, Faculty of Science, The University of Auckland, Auckland, New Zealand. ⁶Psychology Research Laboratory, McLean Hospital, Belmont, Massachusetts, USA. ⁷Department of Psychiatry, Harvard Medical School, Boston, Massachusetts, USA. ⁸Weisskopf Child Evaluation Center, Department of Pediatrics, University of Louisville, Louisville, Kentucky, USA. ⁹Division of Child Neurology, Department of Neurology, University of Louisville, School of Medicine, Louisville, Kentucky, USA. ¹⁰Department of Medical and Molecular Genetics, School of Medicine, and ¹¹Department of Neurology, Division of Pediatric Neurology, School of Medicine, Indiana University, Indianapolis, Indiana, USA. ¹²Division of Genetics, Maine Medical Partners Pediatric Specialty Care, Maine Medical Center, Portland, Maine, USA. ¹³Division of Medical Genetics, Duke University Medical Center, Durham, North Carolina, USA. ¹⁴Department of Pediatrics, University of Toledo Medical College and Northwest Ohio Regional Genetics Center, Toledo, Ohio, USA. ¹⁵Medical Genetics, Shodair Children's Hospital, Helena, Montana, USA. ¹⁶Division of Clinical Genetics, Dartmouth-Hitchcock Medical Center, Lebanon, New Hampshire, USA. ¹⁷Magee-Womens Hospital of University of Pittsburgh Medical Center, Pittsburgh, Pennsylvania, USA. ¹⁸Medical Genetics and Neurodevelopmental Center, St. Vincent Children's Hospital, Indianapolis, Indiana, USA. ¹⁹Division of Medical Genetics, University of Missouri, Columbia, Missouri, USA. ²⁰Geisinger Medical Center, Danville, Pennsylvania, USA. ²¹Division of Genetics, Department of Pediatrics, Louisiana State University Health Sciences Center and Children's Hospital, New Orleans, Louisiana, USA. ²²Department of Pediatrics and Genetics, University of Pennsylvania, and the Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, USA. ²³South Australian Clinical Genetics Service, South Australian Pathology at Women's and Children's Hospital, Adelaide, Australia. ²⁴Department of Paediatrics, The University of Adelaide, Adelaide, Australia. ²⁵Genetics and Molecular Pathology, and South Australian Pathology at Women's and Children's Hospital, Adelaide, Australia. ²⁶Oasi Institute for Research and Care in Mental Retardation and Brain Aging, Troina, Italy. ²⁷VA Boston Healthcare System, Brockton, Massachusetts, USA. ²⁸Departments of Psychiatry and Cellular and Molecular Medicine, University of California, San Diego, La Jolla, California, USA. ²⁹Howard Hughes Medical Institute, University of Washington, Seattle, Washington, USA. ³⁰These authors contributed equally to this work. Correspondence should be addressed to E.E.E. (eee@gs.washington.edu).

ONLINE METHODS

Subjects. We obtained CNV data for 16p12.1 microdeletion analysis on 14,454 subjects comprising two phenotypically distinct cohorts from four independent sources: (i) DNA samples ($n = 11,393$) with indications primarily of intellectual disability/developmental delay and congenital malformation (ID/DD cohort) submitted to Signature Genomic Laboratories during the period of 2007–2008 for CNV analysis (**Supplementary Note**); (ii) CNV data from individuals diagnosed with schizophrenia ($n = 416$) analyzed on the NimbleGen HD2 array-based CGH platform at Cold Spring Harbor Laboratories; (iii) 2,645 DNA samples from individuals with schizophrenia analyzed on the Affymetrix 6.0 platform by the Genetic Association Information Network (GAIN) project for the study of schizophrenia (phs000021.v2.p1); and (iv) 480 individuals from Italy and Australia diagnosed with neurodevelopmental anomalies for CNV analysis using a custom targeted ‘hotspot’ NimbleGen array (**Supplementary Note**). For the replication study, we analyzed CNV data from DNA from 9,254 individuals with ID/DD submitted to Signature Genomic Laboratories in 2009. In addition, we included 96 individuals affected with neurocognitive features and schizophrenia from 26 multiplex families. These families were interviewed, diagnosed and sampled as described^{38,39}. Informed consent was also obtained from a subset of individuals with 16p12.1 microdeletions to perform sequencing analysis of candidate genes within the 16p12.1 microdeletion region. Phenotypic information about probands’ parents was obtained on the basis of family history information gathered during the initial clinic visit before genetic testing (**Supplementary Note**).

Copy-number variation data on controls ($n = 8,540$) consisted of six sets: (i) 671 individuals of European descent with no family history or first-degree relative with amyotrophic lateral sclerosis, ataxia, autism, brain aneurysm, dystonia, Parkinson’s disease or schizophrenia; (ii) 936 middle-aged (40–70 years) individuals of European descent living in the United States tested for statin response and cholesterol levels; (iii) 886 individuals from the Human Genome Diversity Panel (HGDP)¹⁵; (iv) control individuals ($n = 3,181$) used for a large study of schizophrenia³⁵; (v) 446 schizophrenia control samples; and (vi) 2,420 GAIN controls used for a genome-wide association study of schizophrenia. The GAIN cohort was collected to represent unrelated cases or controls. Their health and ethnicity were evaluated through a web-based questionnaire and ascertained by source of sample, geographic representativeness, comparability of cases and controls, and comprehensiveness of trait and phenotypic definitions⁴⁰. For the replication study, we used published CNV data on 2,792 individuals (2,792 of 2,998 passed QC) from Wellcome Trust Case Control Consortium (WTCCC) controls^{41,42} and CNV data from 2,026 individuals excluded for neurological disorders⁴³. The University of Washington Committee on Research Involving Human Subjects and the Institutional Review Board approved this study.

Despite the platform heterogeneity in CNV detection, we and others have shown that these platforms have comparable sensitivity and specificity for large CNV events (>500 kb)¹⁵. We compared the frequency of probands with two large CNVs to that observed in the 2,493 controls¹⁵ (**Supplementary Note**). We used 2,493 control samples from set i ($n = 671$), set ii ($n = 936$) and set iii ($n = 886$), described above, from the discovery set of controls for this purpose. Less than 1% (9/2,493) of our controls had two or more events in excess of 500 kb. As the selection of 16p12.1 deletion probands requires that there is at least one hit already present, we considered only those control individuals who harbored at least one large (>500 kb) CNV as a comparison set. Among the 2,493 controls, 217 individuals have at least one event >500 kb, and of these 217 individuals, only 9 have double hits, for a general population frequency of 4.1% (**Table 3**). We assessed the frequency of two hits in the replication set of 22 cases with the 16p12.1 microdeletions and compared this to the replication set controls ($n = 2,792$) where individual sample identifiers

were available for controls⁴². In the replication cohort, we identified seven cases with a second CNV. Four (18%) of these cases carried a second hit greater than 500 kb in length. We also analyzed the WTCCC data for the second hit⁴². No >500-kb second hit was identified in the three WTCCC controls with the 16p12.1 microdeletion.

Copy-number variation detection. Microarray-based comparative genomic hybridization was performed with a whole-genome bacterial artificial chromosome (BAC) microarray chip (SignatureChipWG), an oligo-based (SignatureChipOS) chip (Agilent Technologies) or NimbleGen ‘hotspot’ array and was validated by FISH (**Supplementary Note**)⁴⁴. To refine the breakpoints of the 16p12.1 deletions identified by whole-genome BAC/oligo arrays, we used a custom, high-density oligonucleotide array (NimbleGen) (**Supplementary Note**). All high-density microarray hybridization experiments were performed as described⁴⁵ using a single, unaffected male (GM15724, Coriell) as reference. For the schizophrenia cohort, DNA samples were evaluated for large CNVs (>100 kb) with whole-genome NimbleGen HD2 arrays, Affymetrix 6.0, or Representational Oligonucleotide Microarray Analysis (ROMA)⁴⁶. The replication set controls were analyzed using Illumina Human Hap550 chip⁴³, Illumina Quad61 or using Affymetrix GeneChip 500K⁴².

16p12 genome structure analysis. Metaphase spreads were obtained from a HapMap lymphoblast cell line (Coriell Cell Repository). FISH experiments were performed using fosmid clones directly labeled by nick-translation with Cy3-dUTP (Perkin-Elmer), Cy5-dUTP (Perkin-Elmer) and fluorescein-dUTP (Enzo). Briefly, 300 ng of labeled probe were used for the FISH experiments; hybridization was performed at 37 °C in 2 × SSC, 50% (v/v) formamide, 10% (w/v) dextran sulfate and 3 μg sonicated salmon sperm DNA in a volume of 10 μl. Post-hybridization washing was at 60 °C in 0.1 × SSC (three times, high stringency). Nuclei were simultaneously DAPI stained. Digital images were obtained using a Leica DMRXA2 epifluorescence microscope equipped with a cooled CCD camera (Princeton Instruments). DAPI, Cy3, Cy5 and fluorescein fluorescence signals, detected with specific filters, were recorded separately as grayscale images. Pseudo-coloring and merging of images were performed using Adobe Photoshop software. A minimum of 50 interphase cells were scored for each experiment.

38. DeLisi, L.E. *et al.* Clinical characteristics of schizophrenia in multiply affected Spanish origin families from Costa Rica. *Psychiatr. Genet.* **11**, 145–152 (2001).
39. DeLisi, L.E. *et al.* Genome-wide scan for linkage to schizophrenia in a Spanish-origin cohort from Costa Rica. *Am. J. Med. Genet.* **114**, 497–508 (2002).
40. Manolio, T.A. *et al.* New models of collaboration in genome-wide association studies: the Genetic Association Information Network. *Nat. Genet.* **39**, 1045–1051 (2007).
41. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447**, 661–678 (2007).
42. Kirov, G. *et al.* Support for the involvement of large copy number variants in the pathogenesis of schizophrenia. *Hum. Mol. Genet.* **18**, 1497–1503 (2009).
43. Shaikh, T.H. *et al.* High-resolution mapping and analysis of copy number variations in the human genome: a data resource for clinical and research applications. *Genome Res.* **19**, 1682–1690 (2009).
44. Bejjani, B.A. *et al.* Use of targeted array-based CGH for the clinical diagnosis of chromosomal imbalance: is less more? *Am. J. Med. Genet. A.* **134**, 259–267 (2005).
45. Selzer, R.R. *et al.* Analysis of chromosome breakpoints in neuroblastoma at sub-kilobase resolution using fine-tiling oligonucleotide array CGH. *Genes Chromosom. Cancer* **44**, 305–319 (2005).
46. Sebat, J. *et al.* Strong association of *de novo* copy number mutations with autism. *Science* **316**, 445–449 (2007).