Published in final edited form as:
N Engl J Med. 2012 October 4; 367(14): 1321–1331. doi:10.1056/NEJMoa1200395.

Phenotypic Heterogeneity of Genomic Disorders and Rare Copy-Number Variants

Santhosh Girirajan, M.B., B.S., Ph.D., **Jill A. Rosenfeld, M.S.**, **Bradley P. Coe, Ph.D.**, **Sumit Parikh, M.D.**, **Neil Friedman, M.B., Ch.B.**, **Amy Goldstein, M.D.**, **Robyn A. Filipink, M.D.**, **Juliann S. McConnell, M.S.**, **Brad Angle, M.D.**, **Wendy S. Meschino, M.D.**, **Marjan M. Nezarati, M.D.**, **Alexander Asamoah, M.D.**, **Kelly E. Jackson, M.S.**, **Gordon C. Gowans, M.D.**, **Judith A. Martin, M.D.**, **Erin P. Carmany, M.S.**, **David W. Stockton, M.D.**, **Rhonda E. Schnur, M.D.**, **Lynette S. Penney, M.D.**, **Donna M. Martin, M.D., Ph.D.**, **Salmo Raskin, Ph.D.**, **Kathleen Leppig, M.D.**, **Heidi Thiese, M.S.**, **Rosemarie Smith, M.D.**, **Erika Aberg, M.S.**, **Dmitriy M. Niyazov, M.D.**, **Luis F. Escobar, M.D.**, **Dima El-Khechen, M.S.**, **Kisha D. Johnson, M.S.**, **Robert R. Lebel, M.D.**, **Kiana Siefkas, M.S.**, **Susie Ball, M.S.**, **Natasha Shur, M.D.**, **Marianne McGuire, M.S.**, **Campbell K. Brasington, M.S.**, **J. Edward Spence, M.D.**, **Laura S. Martin, M.D.**, **Carol Clericuzio, M.D.**, **Blake C. Ballif, Ph.D.**, **Lisa G. Shaffer, Ph.D.**, and **Evan E. Eichler, Ph.D.**

Abstract

BACKGROUND—Some copy-number variants are associated with genomic disorders with extreme phenotypic heterogeneity. The cause of this variation is unknown, which presents challenges in genetic diagnosis, counseling, and management.

METHODS—We analyzed the genomes of 2312 children known to carry a copy-number variant associated with intellectual disability and congenital abnormalities, using array comparative genomic hybridization.

RESULTS—Among the affected children, 10.1% carried a second large copy-number variant in addition to the primary genetic lesion. We identified seven genomic disorders, each defined by a specific copy-number variant, in which the affected children were more likely to carry multiple copy-number variants than were controls. We found that syndromic disorders could be distinguished from those with extreme phenotypic heterogeneity on the basis of the total number of copy-number variants and whether the variants are inherited or de novo. Children who carried two large copy-number variants of unknown clinical significance were eight times as likely to have developmental delay as were controls (odds ratio, 8.16; 95% confidence interval, 5.33 to 13.07; $P = 2.11\times10^{-38}$). Among affected children, inherited copy-number variants tended to co-occur with a second-site large copy-number variant (Spearman correlation coefficient, 0.66; P<0.001). Boys were more likely than girls to have disorders of phenotypic heterogeneity (P<0.001), and mothers were more likely than fathers to transmit second-site copy-number variants to their offspring (P = 0.02).

CONCLUSIONS—Multiple, large copy-number variants, including those of unknown pathogenic significance, compound to result in a severe clinical presentation, and secondary copy-

Address reprint requests to Dr. Eichler at the Howard Hughes Medical Institute, Department of Genome Sciences, University of Washington School of Medicine, Foege S-413A, Box 355065, 3720 15th Ave. NE, Seattle, WA 98195, or at eee@gs.washington.edu.
The authors' affiliations are listed in the Appendix.
Dr. Girirajan and Ms. Rosenfeld contributed equally to this article.

number variants are preferentially transmitted from maternal carriers. (Funded by the Simons Foundation Autism Research Initiative and the National Institutes of Health.)

Genomic rearrangements are an important source of genetic and phenotypic variation. Rare, recurrent copy-number variants of pathogenic significance, termed genomic disorders, were originally identified in persons with a characteristic set of clinically recognizable features, such as the Smith–Magenis syndrome, the Sotos syndrome, and the Williams–Beuren syndrome. Although unexplained phenotypic variation and differences in severity have long been recognized among patients with the same genomic disorder,[1–5] comparatively recent discoveries of potentially pathogenic copy-number variants have broadened the phenotypic range associated with a given variant to include entirely distinct diseases. High-throughput analyses of patient populations have implicated the same copy-number variants in diseases, such as schizophrenia,[6] autism,[7] cardiac disease,[8] epilepsy,[9] and intellectual disability.[10] For example, a recurrent deletion on chromosome 15q13.3 has been associated with intellectual disability,[11] schizophrenia,[12] autism,[13] and 1% of idiopathic generalized epilepsy.[14] Similarly, a deletion on chromosome 16p11.2 has been associated with intellectual disability,[15] obesity,[16] schizophrenia,[17] and 1% of sporadic cases of autism.[18] The factors underlying the phenotypic variation associated with seemingly identical genomic alterations have not been entirely clear and present challenges for clinical diagnosis, counseling, and management. Although such copy-number variants confer a risk of disease, they may not be sufficient by themselves to lead to a specific disease outcome, fueling speculation that additional risk factors may account for the variation.[19,20]

We recently proposed a “two-hit,” or second-site, model that is based on the observation that affected persons with a microdeletion on chromosome 16p12.1 are more likely to have additional large copy-number variants than are controls.[21] Our data supported an oligogenic basis, in which the compound effect of a relatively small number of rare variants of large effect contributes to the heterogeneity of genomic disorders, and provided testable predictions of the cause of syndromic disorders and those with phenotypic variation. In the current study, we tested the generalizability of this second-site model by analyzing the genomic context of 72 large, rare copy-number variants known to be associated with a genomic disorder or potentially associated with disease.[22,23] We also examined the relationship between phenotypic severity and the total size and number of copy-number variants.

METHODS

STUDY SAMPLES

We analyzed 32,587 samples from children who had developmental delay with or without congenital malformations; the samples were submitted to Signature Genomic Laboratories from 2008 through 2010. (Details are provided in the Methods section in the Supplementary Appendix, available with the full text of this article at NEJM.org.) Parents or guardians provided written informed consent, or deidentified data were supplied by clinicians according to a protocol approved by the institutional review board.

We performed microarray-based comparative genomic hybridization (array CGH) with a whole-genome bacterial-artificial-chromosome microarray (SignatureChipWG) for 9207 samples and an oligonucleotide-based microarray (Signature-ChipOS, custom-designed by Signature Genomic Laboratories and manufactured by Agilent Technologies or Roche NimbleGen) for 23,380 samples.[24–26] We obtained data on copy-number variation (calls) from 8329 persons who had been found to have no overt neurologic disorders during screening for other studies (controls) (see the Methods section in the Supplementary Appendix). We compared copy-number variants in the 32,587 samples from children who

had developmental delay with those in the 8329 control samples; copy-number variants with overlap of 50% or more of their length were considered to be the same. To assess copy-number variants for a specific phenotype, we examined children with sporadic autism as part of the Simons Simplex Collection. In this analysis, we generated calls for 841 probands, 1651 parents, and 793 siblings from Illumina 1M and 1M Duo arrays, using the same algorithm that was used for control data with respect to copy-number variants.[22,27,28]

DEFINITION OF COPY-NUMBER VARIANTS

We analyzed 72 regions of rare copy-number variation (primary events) that were previously known to be associated with neurodevelopmental phenotypes or a genomic disorder and placed them into two broad groups on the basis of the presence or absence of associated syndromic features (Tables S1, S2, and S3 and Fig. S1 through S4 in the Supplementary Appendix). We designated the genomic disorders on the basis of the cytoband location, followed by the candidate-gene symbol in parenthesis to provide a quick landmark.

We defined second-site copy-number variants on the basis of the following criteria: the copy-number variant exceeded 500 kb, mapped to a genomic location that differed from that of the first-site copy-number variant (i.e., nonallelic), and was apparently unrelated to the first copy-number variant (a criterion that excluded samples carrying unbalanced translocations and other complex rearrangements from this analysis), with a prevalence of less than 0.1% of that in the general population (<8 of the 8329 controls) in order to exclude potential polymorphic loci that might confound our interpretation of enrichment of additional copy-number variants. Our previous analysis of a large set of cases and controls[22,29] supports the fourth criterion, which we did not apply to copy-number variants that were known to be associated with a neurodevelopmental disorder. To assess the coexistence of large variants of unknown significance, we excluded all samples with a copy-number variant that was known to confer risk or its reciprocal copy-number variant, which is a variant resulting from a reciprocal recombination product at a particular disease-associated genomic region (e.g., the reciprocal copy-number variant of a 17p11.2 deletion [resulting in the Smith–Magenis syndrome] is the 17p11.2 duplication [resulting in the Potocki–Lupski syndrome], and vice versa). We did not consider copy-number variants that were called with a relative paucity of probes (for details, see the Methods section in the Supplementary Appendix).

STATISTICAL ANALYSIS

We performed all analyses using nonparametric tests. Statistics were calculated exclusively for autosomal second-site copy-number variants because of the lack of control data on sex chromosomes. The significance of enrichments for second-site copy-number variants was determined with the use of Fisher's exact test on a two-by-two contingency table. Owing to the strong probability that a large copy-number variant would be pathogenic,[22,29] we report nominal P values for specific second-site loci.

RESULTS

SAMPLE ANALYSES

We analyzed 32,587 samples from children with or without congenital malformations that had been submitted to Signature Genomic Laboratories for array CGH analysis (see the Methods section and Fig. S5 and S6 in the Supplementary Appendix). These samples originated from referral centers located primarily throughout the United States. We found a precision (the percentage of true positives in all initially positive results) of more than 0.945 for the discovery of large copy-number variants on the basis of a previous validation of the

data.[22] For children who carried multiple copy-number variants, we validated the findings pertaining to these variants using fluorescent in situ hybridization (FISH, for 213 variants) or array CGH (for 265 variants). We considered an additional 502 variants to be validated because a parent carried the same variant (Table S4 in the Supplementary Appendix). We obtained control calls from samples obtained from 8329 persons who had been screened for overt neurologic disorders in multiple studies.[22]

We focused on 2312 children who were determined to have 1 of 72 primary-site copy-number variants. These included 23 large, rare, pathogenic copy-number variants that had previously been associated with a typical constellation of clinical features (syndromic disorders), 40 rare copy-number variants that had previously been associated with phenotypic heterogeneity,[20,22] and 9 large, rare copy-number variants of uncertain pathogenic significance (Fig. S5 and S6 in the Supplementary Appendix). These 72 copy-number variants (39 deletions and 33 reciprocal duplications) mapped to 39 distinct genomic regions and were associated with clinical features involving various organ systems, including developmental delay or intellectual disability, autism, cardiac abnormalities, speech deficits, craniofacial features, and other previously defined congenital malformations.[22]

ANALYSIS OF PARENTAL DATA

We determined inheritance using the parental data that were available for 653 probands, who collectively carried 66 of the 72 primary copy-number variants (Fig. S7 in the Supplementary Appendix). Of these 66 primary copy-number variants, there were 18 (found in 64 probands) in which the events were exclusively de novo (Fig. 1). Some of these de novo variants are known to cause clinically well-defined syndromes, such as the Smith–Magenis syndrome, the Williams–Beuren syndrome, the Sotos syndrome, and 17q21.31 (*MAPT*) deletion syndrome (Fig. S8 in the Supplementary Appendix). Each of the remaining variants (48 of 66) was inherited in at least one instance, and the extent to which they were inherited (vs. de novo occurrence) ranged from 5.6% to 100%. This second set included copy-number variants known to be associated with a much more variable set of outcomes, ranging from neuropsychiatric disease to severe intellectual disability — such as 16p12.1 (*CDR2*) deletion, 3q29 (*DLG1*) duplication, 15q11.2 (*NIPA2*) deletion, 22q11.2 (*TBX1*) duplication, 15q13.3 (*CHRNA7*) deletion, and 16p11.2 (*TBX6*) duplication[20] — and rare copy-number variants that are potentially pathogenic, although their pathogenicity has not been proved, such as 6q16 (*SIM1*) duplication, 15q13.3 BP4–BP5 duplication, and 15q24 (*PTPN9*) duplication.

SEX BIAS IN GENOMIC DISORDERS

On the basis of the partition between syndromic and phenotypically variable genomic disorders, we tested for a sex bias (see the Methods section in the Supplementary Appendix). We found a significant male bias among children with genomic disorders characterized by phenotypic variation, as compared with children whose disorders were associated with syndromic features ($P<0.001$ by the Mann–Whitney test) (Fig. 2A). This bias was observed with or without consideration of the presence of additional large copy-number variants (Fig. S9 in the Supplementary Appendix).

ADDITIONAL VARIANTS

Next, we tested for the presence of additional large copy-number variants at a second site (Fig. S10 and S11 in the Supplementary Appendix). Of the 2312 affected children known to carry a primary variant, 200 (8.7%; 87 girls and 113 boys) carried at least 1 additional large variant affecting an autosome. We observed a total of 211 second-site variants (median size, 1.37 Mb), with similar numbers of duplications and deletions (108 and 103, respectively)

(Table S5 in the Supplementary Appendix). In cases in which the second-site variant was also previously associated with a genomic disorder (45 of 200, or 22.5%), the rarer variant was considered to be the primary-site variant,[22] since its rarity suggested that it was more likely to have a severe effect than the more common variant (Table S6 and Fig. S10 and S12 in the Supplementary Appendix). Similarly, a third-site large variant, resulting in an overall median variant burden of 2.5 Mb, was observed in 11 of 200 affected children (5.5%) (Table S7 in the Supplementary Appendix).

We detected a significant enrichment of second-site variants in samples from children with a phenotypically variable genomic disorder, as compared with second-site variants in samples from those with a syndromic genomic disorder (156 of 1509, or 10.3%, vs. 44 of 857, or 5.1%; odds ratio, 2.13; 95% confidence interval, 1.5 to 3.08; $P = 4.49\times10^{-6}$) (Fig. S13 and S14 and Table S8 in the Supplementary Appendix).

VARIANTS OF UNKNOWN SIGNIFICANCE

On the basis of these results, we expanded our analysis to determine whether a second-site model for copy-number variants (i.e., the increased clinical effect of two or more independent large variants) also applies to variants of unknown significance. To assess the significance of the occurrence of multiple copy-number variants in affected children, we compared the prevalence of two or more large variants among the case samples with the prevalence among the control samples. As with our ascertainment of case samples, we conditioned the control samples first according to the presence or absence of a large primary-site variant and then according to the size (>300 kb or >500 kb) and type (deletion or duplication) of identified variants. Once again, we considered only autosomal variants. We excluded all case and control samples with variants known to cause a genomic disorder (including reciprocal events) and analyzed only samples from children with rare variants (<0.1% in controls). Strikingly, the presence of two variants exceeding 500 kb and of unknown significance was eight times as likely to occur in a sample from a child with developmental delay or intellectual disability as in a control sample (odds ratio, 8.16; $P = 2.11\times10^{-38}$). This enrichment remained significant even after conditioning for the presence of at least one variant of unknown significance (and exceeding 500 kb) in case and control samples (odds ratio, 3.53; $P = 2.9\times10^{-11}$), which suggests that children with two or more large and rare variants were much more likely to be affected (see the Methods section in the Supplementary Appendix).

ENRICHMENT FOR SECOND-SITE VARIANTS

To further investigate whether second-site variants are a distinctive feature of a subset of phenotypically variable genomic disorders, we stratified our findings in the case samples according to specific primary variants and observed a nominally significant ($P<0.05$) enrichment for the presence of at least one additional large variant at a second site for 7 of 72 genomic disorders (as defined by the primary variant), as compared with control samples, including the 15q11.2 (*NIPA2*) deletion,[12,30] 16p11.2 (*TBX6*) duplication,[15,17] 16p12.1 (*CDR2*) deletion,[21] 16p11.2 (*SH2B1*) distal duplication, 3q29 (*DLG1*) duplication,[24] 17p13.3 (*YWHAE*) duplication,[31] and 15q23q24 (*ETFA*) deletion.[1] Repeating the analysis after removal of very large copy-number variants (>30 Mb) still resulted in significance for the enrichment of second-site variants for the 7 disorders (Fig. S15 and Tables S9 and S10 in the Supplementary Appendix). Given the reduced power of a primary-site–specific analysis (owing to smaller numbers of affected children carrying site-specific variants), we did not expect many P values to remain significant after multiple-testing corrections. However, the most common variant (a 15q11.2 deletion in *NIPA2*, which was found in 166 children) remained significant even under a stringent Bonferroni correction ($P = 0.04$). The prevalence of second-site variants was lowest (≤5%) in children with the Smith–Magenis syndrome, the

Williams–Beuren syndrome, the Sotos syndrome, the Potocki–Lupski syndrome, or the 17q21.31 (*MAPT*) deletion syndrome. The underrepresentation of second-site variants in the syndromic genomic disorders is probably due to negative selection (i.e., persons with syndromic disorders are already severely affected, and an additional imbalance of gene dosage is likely to be incompatible with life).

INHERITANCE STATUS

We investigated the inheritance status of first- and second-site large copy-number variants for 46 children (see the Methods section in the Supplementary Appendix) and determined that 33 of 46 of the second-site variants (72%) were inherited. No parental bias was observed for the primary variants (P = 0.12 by binomial test). However, we observed a significant bias toward maternal inheritance for second-site variants, with 22 maternally inherited variants versus 11 paternally inherited variants (P = 0.02 by binomial test). We next considered patterns of coinheritance between first- and second-site variants — both were de novo in 5 cases (Table S11 in the Supplementary Appendix). In 12 cases, the 2 variants were inherited from the same parent (8 maternal and 4 paternal), whereas in another 12 cases, the 2 variants were inherited from different parents, suggesting no particular bias in coinheritance (P = 0.16 by a binomial test), although our sample size was limited.

CORRELATION BETWEEN INHERITANCE AND SECOND-SITE VARIANTS

To assess our ability to discriminate between syndromic copy-number variants and those with phenotypic variation, we compared the inheritance pattern and prevalence of second-site variants for each child with a first-site variant. We observed a significant increase in both the proportion of inherited first-site variants (P<0.001 by the Mann–Whitney test) and the prevalence of additional variants at a second site (P = 0.02 by Mann–Whitney test) in disorders with phenotypic variation, as compared with syndromic disorders (Table S12 and Fig. S16 in the Supplementary Appendix). In addition, we observed a positive correlation between the inheritance rate of a primary-site variant and the proportion of children carrying a second-site variant (Spearman correlation coefficient, 0.66; P<0.001) (Fig. 2B).

When the first-site variant was primarily de novo (inherited in less than 30% of cases) and the prevalence of a second-site variant in the affected population was less than 10%, the associated disorder was more likely to be classified as syndromic. This observation is consistent with the subjection of syndromic variants to stronger negative selection (i.e., persons carrying the variants rarely reproduce), and such variants are therefore maintained in the population primarily by sporadic mutation. In contrast, variants associated with phenotypic variation are subjected to weaker negative selection. Our data suggest that the prevalence of each type of variant is influenced, to different extents, by the rate of de novo mutation and by the likelihood that a variant will be transmitted to offspring.

CLINICAL EFFECT OF ADDITIONAL VARIANTS

To understand the effect of additional copy-number variants, we examined reported clinical case histories[15,21,32,33] and attempted to gather additional clinical information on 161 children with at least one large variant (96 children carrying only a primary variant and 65 with multiple large variants). A qualitative clinical reassessment of these samples confirmed that among children with the same genomic disorder, those with multiple variants had deficits in more domains than those with a single variant (Table S13 in the Supplementary Appendix). To assess this with more objective criteria, we focused on three of the variants showing an association with the most phenotypically variable disorders and adopted a scoring system based on a checklist of clinical features described for subtelomeric[34] and balanced de novo chromosomal rearrangements[35] (Tables S14 and S15 in the Supplementary Appendix). This checklist comprises overt clinical features that are

discernible during evaluation of a patient, scored on a scale from 1 (few features) to 14 (many features) and thus captures additional phenotypes, although the severity of any particular clinical feature is not measured. The following clinical features were added to the list to account for the breadth of phenotypic heterogeneity observed for some of these variants: neuropsychiatric features such as autism, schizophrenia, attention deficit–hyperactivity disorder, aggressive behaviors, and sleep disturbance; epilepsy or seizures; and specific organ defects.

We scored phenotypes for 16p11.2 (*TBX6*) deletions and duplications, 1q21.1 (*GJA8*) deletions, and 16p12.1 (*CDR2*) deletions and compared the scores of children carrying only these disease-associated copy-number variants with those of children who also carried second-site variants. The phenotypic scores for children with only one variant ranged widely, reflecting the initial broad ascertainment. The scores for children with multiple large variants, however, were consistently higher, indicating an increased prevalence of additional disease features. We observed significantly higher scores for affected children with variants in addition to the 16p11.2 deletion (P = 0.008 by the Mann–Whitney test) and the 1q21.1 deletion (P = 0.006 by the Mann–Whitney test) than for those with a single primary variant (Fig. 3A, and Fig. S17 in the Supplementary Appendix). Among children with the 16p12.1 deletion and another large variant, there was a trend toward increased phenotypic scores (P = 0.13 by the Mann–Whitney test), which suggests that other factors might be contributing to the variation in children with the 16p12.1 deletion, as hypothesized previously.[21]

SIMONS AUTISM STUDY

To assess the effect of additional copy-number variants on a more specific phenotypic feature, we examined samples from a cohort of children with autism, the Simons Simplex Collection, in which detailed, standardized phenotypic assessments had been performed for each proband.[28] We found a strong association between median IQ and the number of genes affected by rare variants (<0.1% in controls and all siblings), which is consistent with earlier observations.[28] The median IQ crossed the threshold for intellectual disability (<70) in probands with 18 or more disrupted genes (P = 0.002 by the Wilcoxon rank-sum test), as compared with samples with fewer than 18 genes (Fig. 3B and 3C). We then analyzed calls from 841 probands and 793 siblings for the presence of two large variants anywhere within the autosomes. Although only 6 children in this set carried two or more large variants, 5 were among the most severely affected, suggesting a correlation between severity and variant burden (Fig. 3B).

DISCUSSION

In our study, we observed considerable variation in the phenotypes associated with several recurrent copy-number variants (specific variants that are relatively prevalent). This finding was complicated by the identification of apparently normal or mildly affected carrier parents with 16p11.2,[15,17,18] 1q21.1,[32] or 16p12.1[21] copy-number variants, suggesting that these variants are critical but not sole determinants of phenotype. Our data are consistent with locus heterogeneity and a modest number of high-impact variants contributing to a spectrum of disease severity within families. The interpretation of variants associated with phenotypic variation remains challenging at the clinical level, but our study provides a step toward understanding factors that contribute to the phenotypic outcome, which may be used for counseling.

Several of our observations are consistent with a simple genetic model: additional copy-number variants or larger variants increase the number of disrupted haplosensitive or triplosensitive genes, resulting in an additive or synergistic effect on neurodevelopmental pathways and disease outcome (Fig. S18 and S19 in the Supplementary Appendix). First,

when only considering the first-site variants, we found that children with syndromic variants had significantly more disrupted genes (as a consequence of their first-site variant) than did those with variants associated with more variable features ($P<0.001$ by the Mann–Whitney test). When we took into account the number of genes affected by additional (second-site and third-site) variants, we observed no significant difference in the total gene-disruption burden between children with syndromic features and those with phenotypic variation ($P = 0.95$ by the Mann–Whitney test). Second, an analysis of the known functions of genes affected by both primary-site and second-site variants suggests that the phenotypic effect of second-site variants is probably due to the disruption of genes that interact, with respect to function, with genes disrupted by the primary-site variant (Table S16 in the Supplementary Appendix).

We propose that a combination of rare and disruptive variants of large effect can contribute to different phenotypic outcomes, including intellectual disability, epilepsy, autism, and schizophrenia. We distinguish primary mutations that sensitize persons to disease from secondary mutational events, which compound at the molecular level to modify the outcome and severity. The mode of inheritance for both primary and secondary mutations, as well as the size and gene content of copy-number variants, is a critical determinant in distinguishing syndromic disorders from mutations with phenotypic variation. Even after excluding known pathogenic variants, we found that children with two or more rare and large variants of unknown significance were eight times as likely to be classified as having developmental delay as were population controls. These data strongly suggest that the overall burden of genes that are affected by large variants may eventually be of prognostic usefulness, allowing clinicians to better anticipate long-term outcomes when the variants are discovered in affected persons. However, caution must be exercised in interpreting these analyses in the context of prenatal testing, since we obtained our data from well-ascertained cases with severe developmental-delay phenotypes. The positive predictive values for disorders enriched for second-site variants range from 0.06 to 0.17, indicating that testing for variants in the context of prenatal evaluation would be of little value. However, positive-likelihood ratios range from 2.6 to 9.0; therefore, we recommend that children for whom there is a clinical suspicion of developmental delay (and thus enriched pretest odds) should be tested for second-site variants (Tables S17 and S18 in the Supplementary Appendix).

With respect to inheritance, we also found a significant bias toward maternal transmission of second-site variants ($P = 0.02$). Independently of this analysis, we found a significant enrichment of boys with variably expressive (but not syndromic) genomic disorders ($P<0.001$). Combined, these results suggest that females are less vulnerable to the effects of large variants than are males. We propose that males, by virtue of carrying a single X chromosome, are already at least partially sensitized because of the inherent exposure to weakly deleterious mutations on the X chromosome. Thus, an average male will require fewer mutational events to cross the threshold to disease. This would explain the long-standing observation of an increased prevalence of neurodevelopmental and neuropsychiatric disease by a factor of 2 to 4 among males as compared with females.[36,37] Females are more likely to transmit secondary variants (because they are less likely to be physically affected by them), whereas males are disproportionately affected by de novo events.[27]

Our study represents a step toward deconvoluting the effect of copy-number variants in disease and understanding, more broadly, the causes of neurologic disease. Our analysis shows that the phenotypic variation of at least seven genomic disorders may be partially explained by the presence of additional large variants. Although we restricted our analysis to large variants, it is likely that other disruptive copy-number variants, such as smaller variants, single-nucleotide changes, and epigenetic or stochastic factors altering the

expression of genes within functionally relevant pathways, also contribute to phenotypic variation.[38]

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Supported by a Simons Foundation Autism Research Initiative award (SFARI 137578, to Dr. Eichler) and a grant from the National Institutes of Health (HD065285, to Dr. Eichler).

We thank the children and their families who supplied the samples; the many clinicians who aided in the evaluation of the children and provided clinical information, including Susan Winter, Stephanie Vallee, Stephanie DeWard, Suneeta Madan-Khetarpal, Jennifer Defant, Janice Zunich, Kimberly Guthrie, Paul Wong, and Anne Hing; and Heather Mefford, Catarina Campbell, Niklas Krumm, James Priest, and Tonia Brown for their critical reading of an earlier draft of the manuscript and technical assistance.

APPENDIX

The authors' affiliations are as follows: the Department of Genome Sciences (S.G., B.P.C., E.E.E.) and Howard Hughes Medical Institute (E.E.E.), University of Washington, Seattle, Signature Genomic Laboratories, PerkinElmer, Spokane (J.A.R., B.C.B., L.G.S.), Providence-Sacred Heart Hospital, Spokane (J.A.M.), Group Health Cooperative, Seattle (K.L., H.T.), and Yakima Valley Memorial Hospital, Yakima (K.S., S.B.) — all in Washington; Cleveland Clinic, Cleveland (S.P., N.F.); the Division of Child Neurology (A.G., R.A.F.) and Department of Pediatrics (J.S.M., M.M.), Children's Hospital of Pittsburgh of UPMC (University of Pittsburgh Medical Center), and the Department of Pediatrics, University of Pittsburgh School of Medicine (M.M.) — both in Pittsburgh; Ann and Robert H. Lurie Children's Hospital (B.A.) and the Department of Pediatrics, Rush University Medical Center (K.D.J.) — both in Chicago; North York General Hospital, Toronto (W.S.M., M.M.N.), the Department of Pediatrics, Dalhousie University, Halifax, NS (L.S.P.), and the Maritime Medical Genetics, Izaak Walton Killam Health Centre, Halifax, NS (E.A.) — all in Canada; Weisskopf Child Evaluation Center, Pediatrics Department, University of Louisville, Louisville, KY (A.A., K.E.J., G.C.G.); Children's Hospital of Michigan, Detroit (E.P.C., D.W.S.), and the Departments of Pediatrics and Human Genetics, University of Michigan Medical Center, Ann Arbor (D.M.M.) — both in Michigan; the Division of Genetics, Department of Pediatrics, Cooper Medical School of Rowan University, Camden, NJ (R.E.S.); the Group for Advanced Molecular Investigation, Graduate Program in Health Sciences, Center for Biological and Health Sciences, Pontifícia Universidade Católica do Paraná and Genetika–Centro de Aconselhamento e Laboratório de Genética — both in Curitiba, Brazil (S.R.); Maine Medical Partners, Pediatric Specialty Care, Portland (R.S.); the Department of Pediatrics, Ochsner Clinic Foundation, New Orleans (D.M.N.); Peyton Manning Children's Hospital at St. Vincent, Indianapolis (L.F.E., D.E.-K.); State University of New York Upstate Medical University, Syracuse (R.R.L.); Rhode Island Hospital–Hasbro Children's Hospital, Providence (N.S.); Department of Pediatrics, Levine Children's Hospital at Carolinas Medical Center, Charlotte, NC (C.K.B., J.E.S.); Nemours Children's Clinic, Jacksonville, FL (L.S.M.); and the Division of Pediatric Genetics, University of New Mexico Health Sciences Center, Division of Clinical Genetics–Dysmorphology, University of New Mexico, Albuquerque (C.C.).

REFERENCES

1. Andrieux J, Dubourg C, Rio M, et al. Genotype-phenotype correlation in four 15q24 deleted patients identified by array-CGH. Am J Med Genet A. 2009; 149A:2813–2819. [PubMed: 19921647]

2. Antonell A, Del Campo M, Magano LF, et al. Partial 7q11.23 deletions further implicate GTF2I and GTF2IRD1 as the main genes responsible for the Williams-Beuren syndrome neurocognitive profile. J Med Genet. 2010; 47:312–320. [PubMed: 19897463]
3. Girirajan S, Vlangos CN, Szomju BB, et al. Genotype-phenotype correlation in Smith-Magenis syndrome: evidence that multiple genes in 17p11.2 contribute to the clinical spectrum. Genet Med. 2006; 8:417–427. [PubMed: 16845274]
4. Liburd N, Ghosh M, Riazuddin S, et al. Novel mutations of MYO15A associated with profound deafness in consanguineous families and moderately severe hearing loss in a patient with Smith-Magenis syndrome. Hum Genet. 2001; 109:535–541. [PubMed: 11735029]
5. Sarasua SM, Dwivedi A, Boccuto L, et al. Association between deletion size and important phenotypes expands the genomic region of interest in Phelan-McDermid syndrome (22q13 deletion syndrome). J Med Genet. 2011; 48:761–766. [PubMed: 21984749]
6. Walsh T, McClellan JM, McCarthy SE, et al. Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. Science. 2008; 320:539–543. [PubMed: 18369103]
7. Sebat J, Lakshmi B, Malhotra D, et al. Strong association of de novo copy number mutations with autism. Science. 2007; 316:445–449. [PubMed: 17363630]
8. Greenway SC, Pereira AC, Lin JC, et al. De novo copy number variants identify new genes and loci in isolated sporadic tetralogy of Fallot. Nat Genet. 2009; 41:931–935. [PubMed: 19597493]
9. Mefford HC, Muhle H, Ostertag P, et al. Genome-wide copy number variation in epilepsy: novel susceptibility loci in idiopathic generalized and focal epilepsies. PLoS Genet. 2010; 6(5) e1000962.
10. Stankiewicz P, Lupski JR. Structural variation in the human genome and its role in disease. Annu Rev Med. 2010; 61:437–455. [PubMed: 20059347]
11. Sharp AJ, Mefford HC, Li K, et al. A recurrent 15q13.3 microdeletion syndrome associated with mental retardation and seizures. Nat Genet. 2008; 40:322–328. [PubMed: 18278044]
12. Stefansson H, Rujescu D, Cichon S, et al. Large recurrent microdeletions associated with schizophrenia. Nature. 2008; 455:232–236. [PubMed: 18668039]
13. Pagnamenta AT, Wing K, Sadighi Akha E, et al. A 15q13.3 microdeletion segregating with autism. Eur J Hum Genet. 2009; 17:687–692. [PubMed: 19050728]
14. Helbig I, Mefford HC, Sharp AJ, et al. 15q13.3 Microdeletions increase risk of idiopathic generalized epilepsy. Nat Genet. 2009; 41:160–162. [PubMed: 19136953]
15. Rosenfeld JA, Coppinger J, Bejjani BA, et al. Speech delays and behavioral problems are the predominant features in individuals with developmental delays and 16p11.2 microdeletions and microduplications. J Neurodev Disord. 2010; 2:26–38. [PubMed: 21731881]
16. Walters RG, Jacquemont S, Valsesia A, et al. A new highly penetrant form of obesity due to deletions on chromosome 16p11.2. Nature. 2010; 463:671–675. [PubMed: 20130649]
17. McCarthy SE, Makarov V, Kirov G, et al. Microduplications of 16p11.2 are associated with schizophrenia. Nat Genet. 2009; 41:1223–1227. [PubMed: 19855392]
18. Weiss LA, Shen Y, Korn JM, et al. Association between microdeletion and microduplication at 16p11.2 and autism. N Engl J Med. 2008; 358:667–675. [PubMed: 18184952]
19. Dipple KM, McCabe ER. Phenotypes of patients with “simple” Mendelian disorders are complex traits: thresholds, modifiers, and systems dynamics. Am J Hum Genet. 2000; 66:1729–1735. [PubMed: 10793008]
20. Girirajan S, Eichler EE. Phenotypic variability and genetic susceptibility to genomic disorders. Hum Mol Genet. 2010; 19:R176–R187. [PubMed: 20807775]
21. Girirajan S, Rosenfeld JA, Cooper GM, et al. A recurrent 16p12.1 microdeletion supports a two-hit model for severe developmental delay. Nat Genet. 2010; 42:203–209. [PubMed: 20154674]
22. Cooper GM, Coe BP, Girirajan S, et al. A copy number variation morbidity map of developmental delay. Nat Genet. 2011; 43:838–846. [PubMed: 21841781]
23. Kaminsky EB, Kaul V, Paschall J, et al. An evidence-based approach to establish the functional and clinical significance of copy number variants in intellectual and developmental disabilities. Genet Med. 2011; 13:777–784. [PubMed: 21844811]

24. Ballif BC, Theisen A, Coppinger J, et al. Expanding the clinical phenotype of the 3q29 microdeletion syndrome and characterization of the reciprocal microduplication. Mol Cytogenet. 2008; 1:8. [PubMed: 18471269]
25. Ballif BC, Theisen A, McDonald-McGinn DM, et al. Identification of a previously unrecognized microdeletion syndrome of 16q11.2q12.2. Clin Genet. 2008; 74:469–475. [PubMed: 18811697]
26. Duker AL, Ballif BC, Bawle EV, et al. Paternally inherited microdeletion at 15q11.2 confirms a significant role for the SNORD116 C/D box snoRNA cluster in Prader-Willi syndrome. Eur J Hum Genet. 2010; 18:1196–1201. [PubMed: 20588305]
27. O'Roak BJ, Vives L, Girirajan S, et al. Exome sequencing in sporadic autism reveals a highly interconnected protein network and extreme locus heterogeneity. Nature. 2012; 485:246–250. [PubMed: 22495309]
28. Sanders SJ, Ercan-Sencicek AG, Hus V, et al. Multiple recurrent de novo CNVs, including duplications of the 7q11.23 Williams syndrome region, are strongly associated with autism. Neuron. 2011; 70:863–885. [PubMed: 21658581]
29. Itsara A, Cooper GM, Baker C, et al. Population analysis of large copy number variants and hotspots of human genetic disease. Am J Hum Genet. 2009; 84:148–161. [PubMed: 19166990]
30. Mefford HC, Cooper GM, Zerr T, et al. A method for rapid, targeted CNV genotyping identifies rare variants associated with neurocognitive disease. Genome Res. 2009; 19:1579–1585. [PubMed: 19506092]
31. Bruno DL, Anderlid BM, Lindstrand A, et al. Further molecular and clinical delineation of co-locating 17p13.3 microdeletions and microduplications that show distinctive phenotypes. J Med Genet. 2010; 47:299–311. [PubMed: 20452996]
32. Mefford HC, Sharp AJ, Baker C, et al. Recurrent rearrangements of chromosome 1q21.1 and variable pediatric phenotypes. N Engl J Med. 2008; 359:1685–1699. [PubMed: 18784092]
33. Rosenfeld JA, Stephens LE, Coppinger J, et al. Deletions flanked by breakpoints 3 and 4 on 15q13 may contribute to abnormal phenotypes. Eur J Hum Genet. 2011; 19:547–554. [PubMed: 21248749]
34. de Vries BB, White SM, Knight SJ, et al. Clinical studies on submicroscopic subtelomeric rearrangements: a checklist. J Med Genet. 2001; 38:145–150. [PubMed: 11238680]
35. Feenstra I, Hanemaaijer N, Sikkema-Raddatz B, et al. Balanced into array: genome-wide array analysis in 54 patients with an apparently balanced de novo chromosome rearrangement and a meta-analysis. Eur J Hum Genet. 2011; 19:1152–1160. [PubMed: 21712853]
36. Leonard H, Wen X. The epidemiology of mental retardation: challenges and opportunities in the new millennium. Ment Retard Dev Disabil Res Rev. 2002; 8:117–134. [PubMed: 12216056]
37. Ropers HH. Genetics of early onset cognitive impairment. Annu Rev Genomics Hum Genet. 2010; 11:161–187. [PubMed: 20822471]
38. Raj A, Rifkin SA, Andersen E, van Oudenaarden A. Variability in gene expression underlies incomplete penetrance. Nature. 2010; 463:913–918. [PubMed: 20164922]

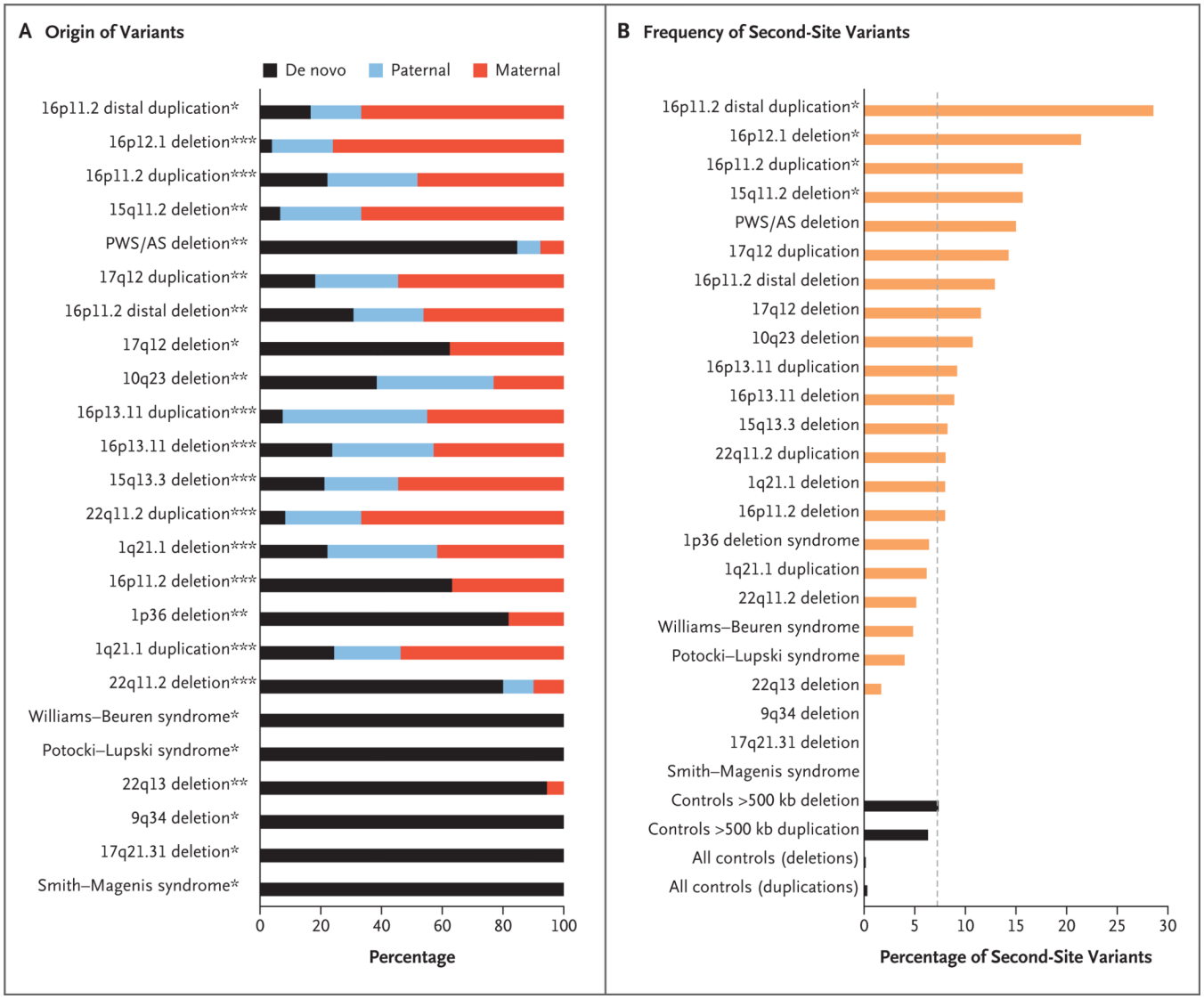

Figure 1. Inheritance Pattern of Copy-Number Variants and Frequency of Second-Site Variants Associated with a Genomic Disorder

In Panel A, the histogram shows the inheritance pattern of copy-number variants. Inheritance information for at least 5 children is shown. The asterisks denote the number of children for whom parental data were available: one asterisk, 5 to 10 children; two asterisks, 11 to 20 children; and three asterisks, more than 20 children. Most copy-number variants that were associated with syndromic disorders arose de novo, whereas the recently discovered variants associated with variable phenotypes were highly inherited. (A complete list of inheritance information for all 72 variants included in this study is provided in Table S12 in the Supplementary Appendix.) AS denotes Angelman syndrome, and PWS Prader–Willi syndrome. In Panel B, the histogram shows the incidence of large secondary copy-number variants in a representative set of variants associated with genomic disorders in samples obtained from 32,587 children with developmental delay or congenital anomalies. Results are displayed in the descending order of frequency of second-site variants. Frequencies of another large variant among controls (conditioned for deletions or duplications of >500 kb) and of 2 large variants (>500 kb) in the general control population (unconditioned) are indicated by black bars. An asterisk indicates a significant enrichment,

as compared with controls who had either a deletion or duplication of the first-site variant. Only four disorders with significant enrichment are represented in this figure. To account for sex bias and the lack of control data on sex chromosomes, these data represent only autosomal second hits.

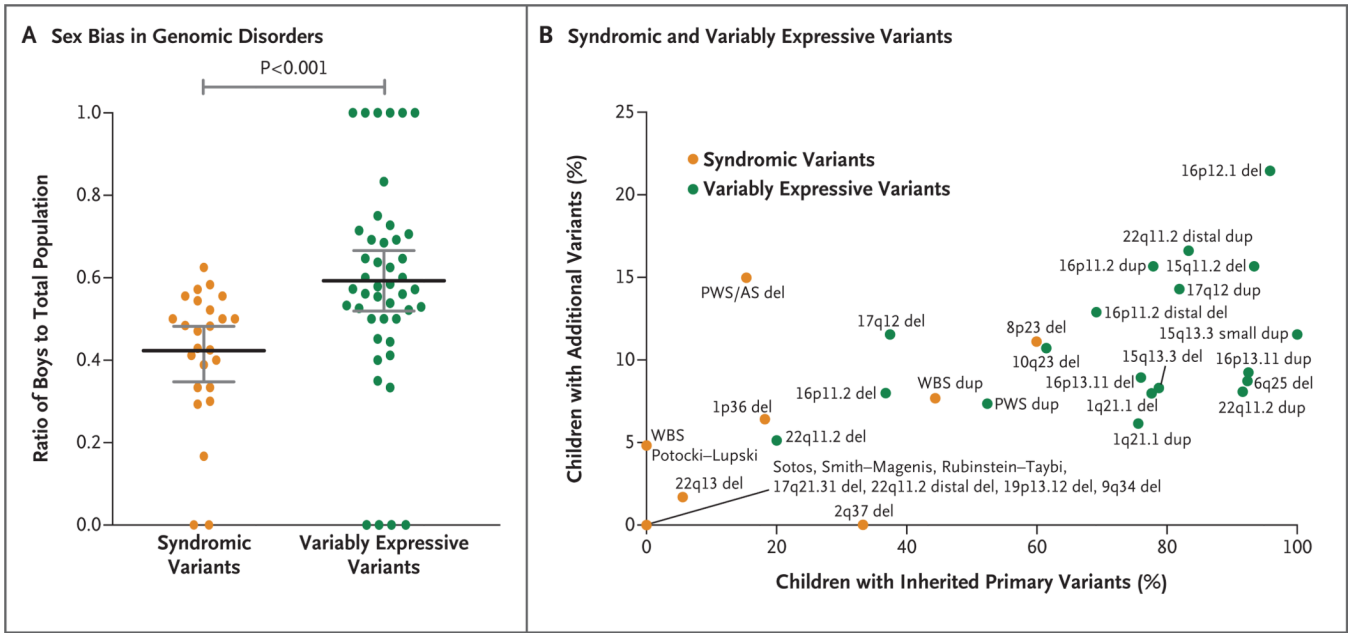

Figure 2. Sex Bias in Genomic Disorders Associated with Phenotypic Variation and the Correlation between the Inheritance of Variants and Incidence of Second-Site Variants

Panel A shows the proportion of boys with all 72 genomic disorders, including syndromic copy-number variants and those with variable features. Boys were more likely than girls to be affected with disorders of phenotypic heterogeneity (P<0.001 by the Mann–Whitney test). Panel B shows the percentage of inherited first-site copy-number variants and the incidence of rare variants at second sites for a representative set of genomic disorders. Each data point represents a genomic disorder. A strong correlation was observed (Spearman correlation coefficient, 0.68; P<0.001) when genomic disorders affecting more than five children were analyzed. The two categories of syndromic disorders and disorders with phenotypic heterogeneity cluster separately according to the percentage of inherited first-site variants. Seven genomic disorders are represented by a single data point at coordinates 0, 0, and two disorders by a single data point at 0, 5. (Additional information regarding variants that are not represented in this figure is provided in Fig. S16 in the Supplementary Appendix.) AS denotes Angelman syndrome, PWS Prader–Willi syndrome, and WBS Williams–Beuren syndrome.

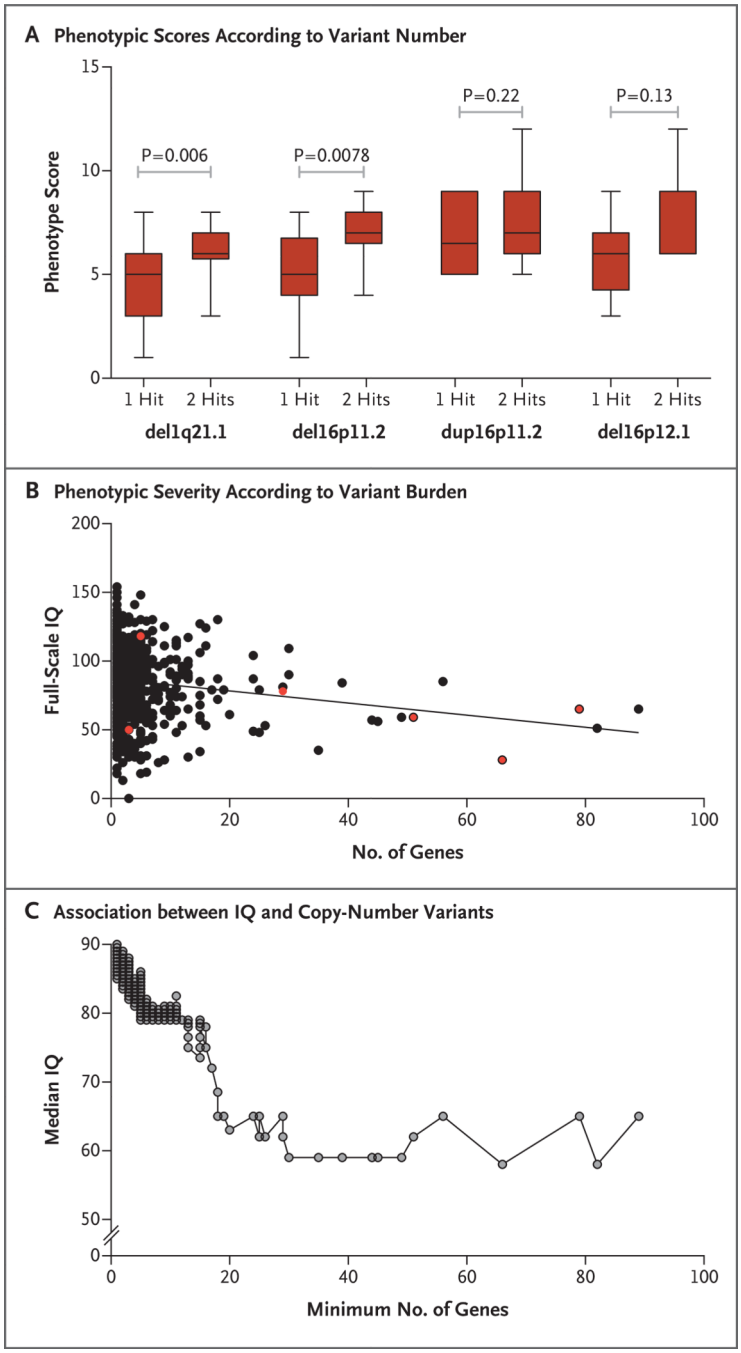

Figure 3. Phenotypic Variation Associated with Additional Large Copy-Number Variants

Panel A shows phenotypic scores for four copy-number variants found in children with developmental delay, according to whether they had a single variant (1 hit) or an additional large variant (2 hits). The scores range from 0 to 14, with higher scores indicating extensive phenotypic heterogeneity. (Details of the scoring system are provided in Tables S13 and S14 in the Supplementary Appendix.) Phenotypic scores for children with multiple large variants were consistently higher than those for children with a single variant, indicating the increased prevalence of additional features of the disorder. The numbers of children with each variant were as follows: 1q21.1 (*GJA8*) deletion: 54 with a single variant and 10 with two large variants; 16p11.2 (*TBX6*) deletion: 16 with a single variant and 13 with two large

variants; 16p11.2 (*TBX6*) duplication: 10 with a single variant and 8 with two large variants; and 16p12.1 (*CDR2*) deletion: 16 with a single variant and 7 with two large variants. The horizontal line within each box represents the median value; the bottom and top lines of the box represent the 25th and 75th percentiles, respectively; and the horizontal lines below and above the box represent the lowest and highest values, respectively. Panel B shows phenotypic severity as an outcome of the burden of rare copy-number variants for autism. The data points represent the analysis of full-scale IQ (y axis) in the context of the number of genes (x axis) disrupted by rare copy-number variants. Red data points represent samples with two large (>500-kb) variants. The trend line running through the data points shows correlation. Panel C shows the association between median IQ and the minimum number of genes disrupted by copy-number variants, indicating a striking reduction in median IQ with an increasing number of affected genes. In probands with 18 or more affected genes, the median full-scale IQ drops below the threshold for intellectual disability (70 points) and is significantly reduced, as compared with probands with fewer than 18 affected genes ($P = 0.002$ by the Wilcoxon rank-sum test).