Rare variants in the genetic background modulate the expressivity of neurodevelopmental disorders

Lucilla Pizzo[1], Matthew Jensen[1], Andrew Polyak[1,2], Jill A. Rosenfeld[3], Katrin Mannik[4,5] Arjun Krishnan[6,7], Elizabeth McCready[8], Olivier Pichon[9], Cedric Le Caignec[9,10], Anke Van Dijck[11], Kate Pope[12], Els Voorhoeve[13], Jieun Yoon[1], Paweł Stankiewicz[3], Sau Wai Cheung[3], Damian Pazuchanics[1], Emily Huber[1], Vijay Kumar[1], Rachel Kember[14], Francesca Mari[15,16], Aurora Curró[15,16], Lucia Castiglia[17], Ornella Galesi[17], Emanuela Avola[17], Teresa Mattina[18], Marco Fichera[18], Luana Mandarà[19], Marie Vincent[9], Mathilde Nizon[9], Sandra Mercier[9], Claire Bénéteau[9], Sophie Blesson[20], Dominique Martin-Coignard[21], Anne-Laure Mosca-Boidron[22], Jean-Hubert Caberg[23], Maja Bucan[14], Susan Zeesman[24], Małgorzata J.M. Nowaczyk[24], Mathilde Lefebvre[25], Laurence Faivre[26], Patrick Callier[22], Cindy Skinner[27], Boris Keren[28], Charles Perrine[28], Paolo Prontera[29], Nathalie Marle[22], Alessandra Renieri[15,16], Alexandre Reymond[4], R. Frank Kooy[11], Bertrand Isidor[9], Charles Schwartz[27], Corrado Romano[17], Erik Sistermans[13], David J. Amor[12], Joris Andrieux[30], and Santhosh Girirajan[1,*]

1. Department of Biochemistry and Molecular Biology. The Pennsylvania State University, University Park, Pennsylvania, USA.
2. St. Georges University School of Medicine, Grenada.
3. Department of Molecular & Human Genetics, Baylor College of Medicine, Houston, Texas, USA.
4. Center for Integrative Genomics, University of Lausanne, Lausanne, Switzerland.
5. Estonian Genome Center, Institute of Genomics, University of Tartu, Tartu, Estonia.
6. Department of Computational Mathematics, Science and Engineering, Michigan State University, East Lansing, Michigan, USA.
7. Department of Biochemistry and Molecular Biology, Michigan State University, East Lansing, Michingan, USA.
8. Department of Pathology and Molecular Medicine, McMaster University, Hamilton, Ontario, Canada.
9. CHU Nantes, Medical genetics department, Nantes, France.
10. INSERM, UMR1238, Bone sarcoma and remodeling of calcified tissue, Nantes, France.
11. Department of Medical Genetics, University and University Hospital Antwerp, Antwerp, Belgium.
12. Murdoch Children's Research Institute and University of Melbourne, Department of Paediatrics, Royal Children's Hospital, Melbourne, Australia.
13. Department of Clinical Genetics, VU University Medical Center Amsterdam, The Netherlands.
14. Department of Genetics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA.
15. Medical Genetics, University of Siena, Siena, Italy.
16. Medical Genetics, Azienda Ospedaliera Universitaria Senese, Siena, Italy.
17. Oasi Research Institute – IRCCS, Troina, Italy.
18. Medical Genetics, University of Catania School of Medicine, Catania, Italy.
19. Medical Genetics, ASP Ragusa, Ragusa, Italy.
20. Department of genetics, Bretonneau university hospital, Tours, France.

21. Service de Cytogénétique, CHU de Le Mans, Le Mans, France.
22. Laboratoire de Génétique Chromosomique et Moléculaire, CHU Dijon, France.
23. Centre Hospitalier Universitaire de Liège. Domaine Universitaire du Sart Tilman, Liège, Belgium.
24. McMaster University, Hamilton, Ontario, Canada.
25. Centre de Génétique. Hôpital d'Enfants Dijon, France.
26. Center for Rare Diseases and Reference Developmental Anomalies and Malformation Syndromes. Dijon, France.
27. Greenwood Genetic Center. Greenwood, South Carolina, USA.
28. Hôpital La Pitié Salpêtrière. Paris, France.
29. Medical Genetics Unit, Hospital "Santa Maria della Misericordia", Perugia, Italy.
30. Institut de Génétique Médicale, Hôpital Jeanne de Flandre, CHRU de Lille, Lille, France.

***Correspondence:**
Santhosh Girirajan, MBBS, PhD
Department of Biochemistry and Molecular Biology
Department of Anthropology
205A Life Sciences Building
Pennsylvania State University
University Park, PA 16802
Phone: 814-865-0674
E-mail: sxg47@psu.edu

ABSTRACT

Rare copy-number variants (CNVs) and gene-disruptive mutations associated with neurodevelopmental disease are characterized by phenotypic heterogeneity. When affected children inherit these mutations, they usually present more severe features than carrier parents, leading to challenges in diagnosis and management. To understand how the genetic background modulates phenotypes of these variants, we analyzed clinical and exome-sequencing data from 757 probands and 233 parents and siblings who carry disease-associated mutations. We found that the number of rare pathogenic secondary mutations in developmental genes (second-hits) modulates the expressivity of disease in probands with 16p12.1 deletion (n=26, p=0.014) and in autism probands with gene-disruptive mutations (n=184, p=0.031) when compared to their carrier family members. Probands with 16p12.1 deletion and a strong family history of neuropsychiatric disease were more likely to manifest multiple and more severe clinical features (p=0.035) and carry a higher burden of second-hits compared to those with mild or no family history (p=0.001). The amount of secondary variants determined the severity of cognitive impairment in 432 probands carrying pathogenic rare CNVs or *de novo* mutations in disease genes and negatively correlated with head size in 84 probands with 16p11.2 deletion, suggesting an effect of the genetic background across multiple phenotypic domains. These second-hits involved known disease genes, such as *SETD5*, *AUTS2*, and *NRXN1*, and novel candidate modifiers, such as *CDH23*, *RYR3*, and *DNAH3*, affecting core developmental processes. Our findings suggest that in the era of personalized medicine, accurate diagnosis will require complete evaluation of the genetic background even after a candidate gene mutation is identified.

INTRODUCTION

Significant advances in high-throughput genomic sequencing technologies have helped to identify hundreds of genes as risk factors for neurodevelopmental and neuropsychiatric disorders, such as autism, intellectual disability, schizophrenia, and epilepsy. For example, in 2002 only 2-3% of autism cases were explained by genetic factors, whereas current studies suggest that rare disruptive mutations, including copy-number variants (CNVs) and single-nucleotide variants (SNVs), account for 10-30% of autism cases[1-5]. Despite initial claims of association with a specific disorder or syndrome, several of these pathogenic variants show incomplete penetrance and variable expressivity[6-9]. For example, the 16p11.2 BP4-BP5 deletion (OMIM #611913)[10] was first described in children with autism, but further studies on other clinical and population cohorts demonstrated that this deletion is also associated with individuals with intellectual disability and developmental delay (ID/DD), obesity, epilepsy, cardiac disease, and scoliosis, and only about 24% of cases manifest an autism phenotype[10-20]. Phenotypic variability is not restricted to multi-genic CNVs but has also been reported for single genes with disruptive mutations, including *DISC1, PTEN, SCN1A, CHD2, NRXN1, FOXP2,* and *GRIN2B* [8,21-28]. While some of these effects could be due to allelic heterogeneity, phenotypic variability among carriers of the same molecular lesion suggests a strong role for variants in the genetic background[28-32]. For example, in a large family described by St. Clair and colleagues, carriers of a balanced translocation disrupting *DISC1* manifested a wide range of neuropsychiatric features, including schizophrenia, bipolar disorder and depression[33-35].

This phenomenon was exemplified by our delineation of a 520-kbp deletion on chromosome 16p12.1 (OMIM #136570) that is associated with developmental delay and extensive phenotypic variability[36]. Interestingly, in most cases, this deletion was inherited from a

parent who also manifested mild neuropsychiatric features, and the severely affected children were more likely to carry another large (>500 kbp) rare CNV. We hypothesized that while each pathogenic primary mutation sensitizes the genome to varying extents, additional secondary variants in the genetic background modulate the ultimate clinical manifestation.

In this study, we evaluated 757 probands and 233 family members carrying disease-associated primary mutations (17 rare CNVs or disruptive mutations in 301 genes). A comparison of the genetic background between probands and parents or siblings showed that in the presence of the same primary mutation, variability and severity of neurodevelopmental disease is contingent upon the number of rare pathogenic variants in the genetic background.

SUBJECTS AND METHODS

Patient recruitment and clinical data ascertainment

We obtained clinical and/or genomic data from 141 children carrying the 16p12.1 deletion, as well as 39 deletion carrier and 30 non-carrier parents. Probands and parents recruited through direct contact provided consent according to the protocol reviewed and approved by The Pennsylvania State University Institutional Review Board. When individuals were not contacted directly, de-identified phenotypic (case histories and clinical information) and genomic data were used; as such, these cases were exempt from IRB review and conformed to the Helsinki Declaration.

We extracted clinical information from medical records or clinical questionnaires completed by different physicians from 180 carrier individuals and available family members. Medical records were comprehensively reviewed for medical history of the probands, including developmental milestones, anthropometric measures, clinical diagnosis of nervous system, cardiac, visual, gastrointestinal, urinary and reproductive organ defects, as well as clinical notes describing tests and observations of cognitive, neurological, and behavioral features. Family history of behavioral, developmental and psychiatric features was also assessed from the medical records. In addition to or in the absence of medical records, information regarding prenatal and developmental history, presence or absence of overt phenotypes such as craniofacial, skeletal and muscular features, and cognitive and behavioral features of the probands were obtained through clinical questionnaires completed by physicians. Clinical questionnaires and direct interviews with parents were also used to collect family history information and history of neuropsychiatric features of the parents, including depression, learning difficulties, alcohol/drug abuse, attention-deficit disorder, bipolar disorder, behavioral issues, delusions, and hallucinations.

We used a modified de Vries scoring system for quantifying the number and severity of phenotypic abnormalities in affected children, which allows for a uniform assessment of developmental phenotypes from clinical records (Table S1)[6,37-40]. Originally used for characterizing phenotypes associated with subtelomeric and balanced chromosomal rearrangements, this method, used reliably in several studies, allows for a uniform assessment of developmental phenotypes from clinical records[38-40]. Using keyword searches for more than 50 clinical terms, we binned specific features into nine broad phenotypic categories, including craniofacial/skeletal features, head phenotype (macro/microcephaly), growth, developmental/speech/motor delay/intellectual disability, abnormal behavior, hypo/hypertonia, epilepsy, congenital malformation, and family history of neurodevelopmental and psychiatric features. Each feature was given a score ranging from 0 (feature not present) to a maximum of 4 (severe feature) based on presence of a specific feature and its severity, and a total score ranging from 1 (few features) to 18 (many severe features) was calculated to denote the number and severity of the phenotypic categories affected in each proband.

Family history information was used to bin families with the deletion into strong, mild or negative family history categories based on the severity of neurodevelopmental or psychiatric features (Figure S1). We considered families to have a strong family history when either parent presented at least one major psychiatric or developmental feature (such as intellectual disability, schizophrenia, bipolar disorder, congenital features, or multiple episodes of epilepsy) or two or more mild psychiatric features (such as mild depression, difficulties in school, or alcohol/drug abuse), and/or siblings exhibited neurodevelopmental or behavioral features (such as developmental/speech delay, intellectual disability, or autism). We considered families to have mild family history when parents presented one mild psychiatric feature. Families were

categorized as having a negative family history when neither parents nor siblings exhibited any of the assessed features.

Genomic and clinical data for families with autism were obtained from the Simons Foundation Autism Research Initiative (SFARI) following appropriate approvals. Full-scale intellectual quotient (FSIQ), Social Responsiveness Scale (SRS) T-scores and body-mass index (BMI) z-scores data were obtained for 53 individuals carrying rare disease-associated CNVs and 295 individuals with *de novo* mutations from the Simons Simplex Collection (SSC)[41,42]. FSIQ, SRS T-scores, BMI and head circumference z-scores were downloaded for 86 individuals from families with 16p11.2 BP4-BP5 deletion collected as part of the Simons Variation in Individuals Project (SVIP).

Exome sequencing and SNP arrays

We generated exome sequencing and SNP microarray data for 105 individuals from 26 families with the 16p12.1 deletion using standardized pipelines[43-45]. Genomic DNA was extracted from peripheral blood using QiaAMP maxi DNA extraction kit (Qiagen) and treated with RNAse. DNA was then quantified using Qubitor PicoGreen methods (Thermo Fisher Scientific), and sample integrity was assessed in agarose gel. After passing quality control, exome sequencing was performed on these samples at the Genomic Services Lab at the HudsonAlpha Institute for Biotechnology (n=57) and at the Genomics Core Facility, The Huck Institutes of the Life Sciences, The Pennsylvania State University (n=48). Genomic libraries were constructed using the NimbleGen SeqCap EZ Exome v3 capture kit, and paired-end sequencing (2×100 bp) was performed using Illumina HiSeq v4. Reads were trimmed using Sickle v.1.33 and aligned using BWA-MEM v.0.7.13 to the 1000 Genomes Project Phase I reference genome (hg19/GRCh37)[46].

Mapped reads were then processed according to the GATK v.3.5 Best Practices Pipeline, including removal of duplicate reads, local realignment of insertion/deletion sites, and recalibration of base quality scores[43-45]. In order to detect splice-site mutations in regions flanking exons, we extended our target regions by 100 bp at the 5' and 3' ends of each exon, increasing our total extended target size to 137 Mbp.

SNVs and small indels located within 100 bp of the exon capture probes were called for individual samples using GATK v.3.5 HaplotypeCaller, and were jointly genotyped using GATK GenotypeGVCFs. After variant quality score recalibration, called variants were annotated using Annovar v.2016Feb01 including predictive tools for pathogenicity of the alternate allele (Mutation Taster, CADD score), allele frequency in the Exome Aggregation Consortium (ExAC database), and Residual Variation Intolerance Scores (RVIS)[47-50]. Called variants were filtered for the following attributes: quality score ≥50, read depth ≥8, number of reads with 0 mapping quality ≤4 and ≤10% of all reads, ratio of quality score to alternate reads ≥1.5, and allele balance between 0.25 and 0.75 (heterozygous) or ≥0.9 (homozygous). An average of 62.6 Mbp representing 99.1 % of the primary target (64 Mbp) was achieved at ≥8X coverage across the 105 samples (excluding padded regions), with an average number of 18,900 variants called per sample (Table S2). Loss-of-function mutations (LoF), including stopgain, frameshift insertion or deletion, splice-site mutations (predicted by MutationTaster as *disease causing*, "D", or *disease causing automatic*, "A"), and *de novo* mutations in probands were visually confirmed basis using Integrative Genomics Viewer (IGV)[51]. Rare (ExAC≤0.1%) missense mutations with Phred-like CADD ≥25 were investigated in sets of genes associated with neurodevelopmental disorders or reported as disease causing in OMIM [2-4,14,41,52-68]. A subset of disruptive mutations in disease-associated genes was validated using Sanger sequencing.

High-resolution microarrays (Illumina Omni 2.5 BeadChip) were performed on 105 individuals (16p12.1 deletion carriers n=59, non-carriers n=46) at the Genomic Services Lab at the HudsonAlpha Institute for Biotechnology (n=38), Yale Center for Genome Analysis (YCGA) (n=43), and the Department of Genome Sciences at the University of Washington (n=24). PennCNV v.1.0.3 was used to identify CNVs from high-resolution array data[69]. Individual and family-based (trios and quads) PennCNV calls were combined for autosomal chromosomes, while CNVs on chromosome X were called only at the individual level. Adjacent CNVs with overlapping base pairs or gaps with <20% of CNV length and <50 kbp were merged. Calls were filtered by size (≥50 kbp in length and containing ≥5 target probes), presence of at least one protein-coding gene (hg19 RefSeq gene), frequency (≤0.1% in a control population of 8,629 individuals[6] as determined by 50% reciprocal overlap), and overlap with segmental duplications and centromere/telomere sequences (≤75%). CNV calls in children and parents were visually confirmed by inspection of log-R ratio (LRR) and B-allele frequency (BAF) plots.

For the autism collection, variant calls (SNVs and CNVs) from 716 individuals from the SSC were obtained from exome and SNP microarray studies[41,42,70]. Variant call files (VCF) and SNP array data from 84 families with 16p11.2 BP4-BP5 deletion were obtained from the SVIP. Single-nucleotide variants and CNVs were obtained and filtered following the same procedures as for the 16p12.1 deletion cohort[43-45,50].

Secondary variant burden analysis

We defined “primary variant” or “first-hit” as follows: (a) rare CNVs previously associated with neurodevelopmental and neuropsychiatric disorders[6], (b) previously reported *de novo* mutations in candidate genes[41,42], and (c) likely damaging variants (loss of function and

missense mutations with CADD≥25) in recurrent genes associated with neurodevelopmental and neuropsychiatric disorders. We defined "secondary variant" or "second-hit" as an additional rare likely pathogenic mutation (<0.1% frequency CNV or SNV with CADD≥25) affecting a functionally intolerant gene in an individual who already carries a disease-associated primary mutation (Figure 1A).

We restricted our search for secondary variants to a subset of genes less likely to harbor mutations in a control population, as a proxy for association of genes with disease. The RVIS has been shown to be a good predictor of gene tolerance to functional variation, and genes within the 20th percentile of RVIS scores have been shown to be enriched in developmental function[42,48]. Secondary variant burden was measured as the number of functionally intolerant genes (with RVIS≤20th percentile) affected in the individual: either carrying rare (frequency in ExAC database ≤0.1%) likely pathogenic variants (loss-of-function and missense mutations with a Phred-like CADD ≥25, representing the top 0.3% of most deleterious mutations in the human genome) or within rare CNVs (found in ≤0.1% of a control population and ≥50 kbp)[47,48]. We did not observe any correlation between the number of secondary variants identified by exome sequencing and the total size (bp) of the region sequenced at ≥8X (Pearson correlation coefficient R= -0.08 for 16p12.1 deletion, p=0.43), which allowed us to directly compare the number of secondary variants within each cohort (Figure S2).

We calculated the frequency of second hits in five subgroups of individuals carrying a disease-associated primary variant (Figure 1B): (a) We analyzed 26 probands, 23 carrier parents and available family members carrying 16p12.1 deletion. (b) We assessed 53 individuals from the SSC cohort who carry rare CNVs associated with syndromic and variably expressive genomic disorders[6]. (c) We analyzed 84 probands and available family members with 16p11.2

BP4-BP5 deletion from the SVIP cohort. (d) We evaluated the frequency of secondary variants in 295 individuals from the SSC cohort reported to carry severe *de novo* loss-of-function mutations in neurodevelopmental genes[41,42]. (e) We assessed 184 affected probands and matched unaffected siblings from the SSC cohort who inherited the same rare (≤0.1% frequency) loss-of-function or likely pathogenic missense primary mutations (CADD ≥25) in genes associated with neurodevelopmental disorders[71-74]

Hierarchical clustering (Ward's method) of the genetic burden in probands, differences in burden between probands and carrier parents, and modified de Vries score was performed using JMP Pro 13.1.0.

Functional analysis of secondary variants

The function of second-hit genes in the 16p12.1 deletion cohort was analyzed using the Ingenuity Pathway Analysis (IPA, Qiagen Bioinformatics) and expression data derived from the GTex database[72]. (a) To identify enrichment in specific canonical pathways among the genes conferring a higher burden in probands, we performed IPA on 219 genes carrying second-site variants in probands and 130 genes in carrier parents, using the Ingenuity Knowledge Base as a reference set (QIAGEN)[75] . Significant enrichment in specific pathways was identified using a one-tailed Fisher's exact test with Benjamini-Hochberg Multiple Testing correction at FDR<0.05 and z-score <-2. (b) Tissue-specific median RPKM expression values for human genes were obtained from the GTex database[72]. A gene was considered to be highly expressed in a specific tissue when its expression was at least two standard deviations greater than the average expression of the gene across 30 tissues, including skeletal muscle, urinary system (kidney and bladder), heart, reproductive system (cervix, vagina, testis, fallopian tube, ovary, prostate,

uterus), digestive system (esophagus, small intestine, colon and stomach), lung and liver. The number of secondary variants in genes with high expression in brain or non-brain tissues (GTex) was compared between probands and carrier parents using a paired t-test.

Gene ontology (GO) enrichment analysis of secondary variants identified in SSC probands with *de novo* mutations and 16p11.2 BP4-BP5 deletion probands from SVIP was performed using the Panther Statistical Overrepresentation test[76]. Biological process GO terms (curated from Panther GO Slim) with significant enrichment for each gene set (FDR<0.05 with Bonferroni correction) were reported. Networks of connected GO terms were created using Cytoscape and the EnrichmentMap plug-in[77,78].

Statistics

Proband-parent or proband-sibling secondary variant burdens and clinical severity scores were compared using a paired t-test. Genetic burden and clinical severity scores between different categories of probands with related or shared primary variants were compared using non-parametric one-tailed Mann-Whitney tests, due to the hypothesis-driven nature of the experiment. Pearson correlation coefficients were calculated between the burden of secondary variants and FSIQ, SRS T-scores, BMI z-scores and head circumference (HC) z-scores. All statistics were calculated using Minitab software.

RESULTS

Secondary variants account for disease expressivity in 16p12.1 deletion probands

We assessed how secondary variants can modulate phenotypes in concert with a first-hit by evaluating 757 affected probands and 233 family members carrying disease-associated primary variants (Figure 1). Using the 16p12.1 deletion as a paradigm for studying the genetic basis of variable expression of disease traits, we first analyzed 180 individuals with this deletion and their non-carrier family members (Figure 1B). The 16p12.1 deletion was inherited in 92.4% of cases, with a significant maternal bias (57.6% maternal (n=53) vs 34.8% paternal (n=32), one-tailed binomial test p=0.02) (Table S3). In accordance with the female protective model described for neurodevelopmental disorders[6,79,80], we observed a significant gender bias among probands with the 16p12.1 deletion (67.9% males versus 32.1% females, one-tailed binomial test p<0.0001). Detailed clinical analysis of 141 affected children with 16p12.1 deletion showed a wide heterogeneity of phenotypes, with a high prevalence of neurodevelopmental, craniofacial and musculoskeletal features (>50%), and variable involvement of other organs and systems (Figure 2A-B, Table S4). In contrast, 32 of 39 (82%) (61.5% females, 38.5% males) carrier parents showed mild cognitive, behavioral and/or psychiatric features (Table S5), consistent with previous reports of cognitive impairment and increased risk for schizophrenia in carriers of the 16p12.1 deletion[7,81,82].

To identify secondary variants within protein-coding regions that contribute to variable phenotypes, we performed exome sequencing and high-resolution SNP arrays in 105 individuals with the 16p12.1 deletion, including 59 deletion carriers (26 probands, 7 carrier siblings, 23 carrier parents, and 3 grandparents) and available non-carrier family members (Table S6). We first evaluated whether the deletion could unmask recessive alleles, and found no rare pathogenic mutations within the seven 16p12.1 genes on the non-deleted chromosome (Table S7). We next

performed a case-by-case analysis of families for second hits elsewhere in the genome by focusing on rare CNVs (<0.1%, ≥50 kbp), *de novo* or rare (ExAC frequency <0.1%) loss-of-function (LoF) mutations, and rare likely pathogenic missense variants (Phred-like CADD ≥25) in disease-associated genes (see Methods). We found that 25/26 probands (96%) carried secondary variants (an average of 2.5 private likely pathogenic mutations per proband) within disease-associated genes, and 4/26 probands (15%) carried another large (≥500 kbp) CNV (Tables S8-S10). For example, in one proband (family GL_001) we identified three disease-associated mutations, including a *de novo* LoF mutation in the intellectual disability-associated gene *SETD5* (c.1623_1624insAC, p.Asp542Thrfs*3) and LoF mutations in *LAMC3* (c.1720C>T, p.Gln574X) and *DMD* (c.9G>A, p.Trp3X), both transmitted from the non-CNV carrying mother[4,83,84] (Figure 2C). Similarly, another proband (family GL_022) carried a stop-gain mutation in the autism-associated gene *OR52M1* (c.264G>A, p.W88X) transmitted from the CNV-carrying mother, and rare LoF mutations in two neuropsychiatric genes, *NALCN* (c.2297C>T, p.Ser766X) and *CCDC137* (c.205G>T, p.Glu69X), transmitted from the father[55,85,86]. Further, in another family (family GL_011), a rare deletion in 2p16.3 inherited from a non-16p12.1 carrier parent, encompassing *NRXN1*, was also identified in a proband. In all of these examples, we found increased variability and more severe clinical features in probands than expected from a single hit alone (Table S11), suggesting that the clinical outcome could be due to a combination of effects due to the second hits and haploinsufficient genes within 16p12.1[4,83,84,87,88].

While private and disease-associated secondary mutations may explain the variable and severe features in the affected children on a case-by-case basis, we lacked the statistical power to implicate individual genes or variants that modulate specific 16p12.1 deletion phenotypes.

Therefore, we performed an integrative global analysis by quantifying rare (frequency ≤0.1%) and likely pathogenic variants (CNVs or SNVs with CADD≥25) within functionally intolerant genes (RVIS≤20th percentile), hereafter referred to as the burden of secondary variants or second hits. Intra-familial comparison of second-hit burden showed that probands have an excess of secondary variants compared to their carrier parents (paired t-test, p=0.014) (Figures 2D, S3 and S4A). Additionally, probands exhibited a greater-number of affected phenotypic domains (p=0.04, modified de Vries scores, Table S1) and showed a trend towards an excess of secondary variants (p=0.09) compared to their carrier siblings (Figures S4B-C). We found an excess of proband second-hit genes that were preferentially expressed in the human brain (p=0.02, Figure S5) and enriched for developmental pathways (Table S12).

The severity and variability of neurodevelopmental features is contingent upon family history of neuropsychiatric disease[80]. In fact, the cognitive and social outcomes in probands with *de novo* 16p11.2 BP4-BP5 deletion or 22q11.2 deletion have been reported to positively correlate with the cognitive and social skills of their parents[89,90]. However, the genetic basis of such background effects has not been sufficiently studied. We assessed the genetic basis of family-specific background effects on the observed inter-familial variability of clinical features in probands with 16p12.1 deletion. We found that probands with a strong family history of neurodevelopmental and psychiatric disease were more likely to present a more severe and heterogeneous clinical presentation than those with mild or negative history, with a higher prevalence of dysmorphic features, epilepsy, and hypotonia (p=0.035) (Figures 3A and 3B). Interestingly, there was not only a higher secondary variant burden in probands with a strong family history (Mann Whitney one-tailed p=0.001), but also a higher difference in burden between probands and their carrier parents (p=0.003) (Figures 3C-3D and Figure S6A-C and

Table S5). Therefore, in families with a strong history of neurodevelopmental and psychiatric disease, severe manifestation of the disorder in the affected proband with 16p12.1 deletion is due to a higher number of rare and potentially pathogenic secondary variants transmitted mostly from non-16p12.1-deletion carrier parents (p=0.03, Figure 3E). Conversely, in families with a mild or negative family history, fewer secondary mutations in functionally intolerant genes are transmitted, contributing to a milder presentation of clinical features (Figure S6A). These results provide further evidence for a role of variants in the genetic background towards both intra- and inter-familial clinical variability observed in families with the 16p12.1 deletion.

Secondary variants modulate quantitative phenotypes among individuals with 16p11.2 deletion and other rare pathogenic CNVs

We assessed whether the second-hit burden modifies quantitative phenotypes in carriers of other CNVs associated with neurodevelopmental phenotypes as well as in autism simplex cases with a previously identified *de novo* mutation (Figure 1B). In probands with disease-associated rare CNVs (n=53, Table S13) from the Simons Simplex Cohort (SSC), we observed a modest but significant negative correlation (Pearson correlation, R=-0.36, p=0.004) between the number of secondary variants and full-scale IQ scores (FSIQ), while no correlation was found for Body-Mass Index (BMI) z-scores and Social Responsiveness Scale (SRS) T-scores, measures for obesity and autism (Figures 4A and S7A-C). Negative correlations between the number of secondary variants and FSIQ held true when we independently analyzed individuals carrying 16p11.2 BP4-BP5 deletion (R=-0.64, p=0.04), 16p11.2 BP4-BP5 duplication (R =-0.74, p=0.007), 1q21.1 duplication (R =-0.60, p=0.14), or 7q11.23 duplication (R =-0.66, p=0.17) (Figure S8). This suggests that each primary mutation sensitizes the genome in a unique way

towards the development of intellectual disability phenotypes, and therefore requires a range of secondary variants that differ in number and effect to lead towards a more severe phenotype (Figure S9).

We further expanded our analysis of secondary variants by evaluating a larger set of 84 families with 16p11.2 BP4-BP5 deletion from the Simons Variation in Individuals Project (SVIP). We observed a higher median number of secondary variants in probands carrying the 16p11.2 deletion that had intellectual disability (FSIQ<70, median=8) compared to those with no intellectual disability (FSIQ≥70, median=7 one-tailed Mann-Whitney, p=0.08, Figure 4B), without a difference in the number of synonymous mutations between the two subgroups (median of 9,957 synonymous changes for FSIQ<70 group versus 10,052 for the FSIQ≥70 group, two-tailed Mann-Whitney, p=0.51, Figure S10A). As previously reported for the burden of second-hit CNVs, carriers of 16p11.2 deletion exhibited a mild negative correlation between FSIQ and secondary variant burden that did not attain statistical significance (Pearson correlation, R=-0.16, p=0.08, Figure S10B)[91]. We hypothesized that this marginal significance compared to that found in 16p11.2 deletion probands from the SSC cohort (Figure S8B) and the previously reported correlation with the second-hit CNV burden could be due to differences in clinical ascertainment. While the SVIP cohort was selected for individuals carrying a 16p11.2 deletion and manifesting a more heterogeneous set of phenotypes, individuals from the SSC cohort were specifically ascertained for idiopathic autism[92]. These differences in ascertainment were also evident by the different distributions of quantitative phenotypes, including BMI, FSIQ, and SRS T-scores in both populations (see Figure S11).

We next assessed whether probands with inherited or *de novo*16p11.2 deletion showed differences in their FSIQ scores. We found that children with an inherited deletion presented

lower FSIQ scores (n=8, median FSIQ=75) than probands with a *de novo* deletion (n=57, median FSIQ=85, one-tailed Mann-Whitney, p=0.006, Figure S12A), consistent with previous reports[19]. This was also consistent with a non-significant increase in the burden of secondary variants among probands with an inherited 16p11.2 deletion compared to those with a *de novo* deletion (one-tailed Mann-Whitney, p=0.06, Figure S12B). Interestingly, the secondary variant burden in 16p11.2 deletion probands negatively correlated with BMI (R=-0.21, p=0.03) and head circumference z-scores (R=-0.27 p=0.008, Figure S13A-B). When we adjusted for age to allow for manifestation of these phenotypes, we still observed a negative correlation for head size (age≥12months, n=80, R=-0.26, p=0.009, Figure 4C) but not for BMI (p>0.05, age>5 years and >10 years) (Figures S13 C-D)[10,19]. The observation that head circumference (HC) z-scores decline steadily (from >2 to <-2 scores) as second hits accumulate suggests that the deletion primarily leads to macrocephaly phenotypes, but the trait is modulated by the presence of second hits[15,19,93]. Our results show that carriers of the 16p11.2 deletion who carry an increased number of secondary variants are more likely to manifest a smaller head size, observed in about 5% of deletion carriers[94,95]. We also note that the second-hit burden did not correlate with SRS T-scores (R=0.00, p=0.5), suggesting a more complex effect of the genetic background towards the penetrance of autism in individuals with the deletion (Figure S14A-B). Overall, these results suggest that in the presence of the 16p11.2 deletion, secondary variants exert a differential role towards variability across multiple phenotypic domains.

Secondary variants modulate disease manifestation among individuals with disruptive mutations in disease-associated genes

We next analyzed 295 simplex cases from the SSC cohort with previously identified *de novo* gene-disruptive mutations within 271 genes, and observed a moderately negative correlation between secondary variant burden and FSIQ scores (Pearson correlation, R=-0.22, p=0.0001, Figure 4D). Within this cohort, individuals with intellectual disability (FSIQ<70, n=93) presented an enrichment of second hits compared to those without intellectual disability (FSIQ ≥70, n=197) (one-tailed Mann-Whitney, p=0.001, Figure S15A)[41,42]. We did not observe a role for secondary variant burden in modulating BMI z-scores (R=0.035, p=0.28, Figure S15B) or SRS T-scores (R=0.10, p=0.05, Figure S15C-D). Interestingly, we found that 17.4% (15/86) of the probands with *de novo* disruptive mutations in recurrent neurodevelopmental genes also carried a rare likely pathogenic secondary variant in other disease-associated genes (Table S14). Moreover, when probands were separated by gender, we observed a higher burden of secondary variants in females compared to males (one-tailed Mann-Whitney, p=0.02, Figure 4E). This supports the hypothesis that females require a higher contribution from the genetic background to reach the genetic threshold for neurodevelopmental disease than males[79,96]. When analyzing disruptive mutations in specific genes, we found that among individuals with damaging mutations in *SCNIA*, the severity of cognitive deficits was contingent upon the number of secondary variants (probands with FSIQ<70, n=8, median=16.5, versus those with FSIQ≥70, n=8, median=8.5, Mann-Whitney, one-tailed p=0.018) (Figure S16). This observation could also explain the diversity of other phenotypes co-occurring with the disruption of the epilepsy-associated *SCN1A* gene, including intellectual disability and autism[26].

While there is a consensus on the pathogenic role of *de novo* gene-disruptive mutations in simplex families, the interpretation of inherited disruptive variants within the same genes is challenging[29,31,32,97,98]. To understand the role of the genetic background in the penetrance of

inherited disruptive mutations in disease-associated genes, we also analyzed 184 pairs of probands and unaffected siblings who inherited the same mutations from a parent (Table S15). We found a greater enrichment of second hits in probands compared to their unaffected siblings (p=0.031, Figure 4F). This result suggests that second hits contribute to increased penetrance of neurodevelopmental phenotypes in children with inherited potentially-pathogenic single-gene mutations.

Secondary variants affect known disease genes and novel modifiers within core developmental processes

To understand how secondary variants modify phenotypes among probands with pathogenic first-hit variants, we performed gene ontology enrichment analysis and characterized the function of second-hit genes from SSC and SVIP probands[76]. Interestingly, we found that genes carrying secondary variants in probands with *de novo* mutations from the SSC cohort are involved in multiple processes, including apoptosis, cell signaling and adhesion, and development (Figure 5A, Table S16). Similarly, second hits found among probands with the 16p11.2 deletion clustered in developmental processes, cell signaling and adhesion, and nervous system processes (Figure 5B, Table S17). These results suggest that secondary variants affect diverse gene categories that consistently involve core cellular and developmental processes. More specifically, when we explored the set of genes carrying secondary single-nucleotide variants in probands with pathogenic CNVs and *de novo* SNVs, we identified 3,197 rare likely-pathogenic mutations encompassing a diverse set of 1,615 functionally intolerant genes. While 854 (53%) genes occurred only once as a second-hit, 761 genes (47%) were identified at least twice in our analysis, with *CDH23*, *RYR3*, *FLNB*, *DNAH3*, *ACF1*, *SPTB* and *HMCN1* observed more than 10

times. Although several of these genes have been individually associated with a disease phenotype, further functional analyses are required to understand their specific contribution towards phenotypic variability[99-104]. Interestingly, 44 of the second-hit genes (such as *CNTNAP2*, *MBD5*, *SCN1A*, *CHD8* and *AUTS2*) have been recurrently associated with neurodevelopmental disorders[71] (Figure S17), and 58 genes have been previously identified as a "causative" gene in simplex autism cases[41,42] (Table S18). We further assessed the location of mutations within a subset of recurrent second-hit genes, including *RIMS1*, *DIP2A*, *KDM5B* and *ACOX2*, and found no specificity for the location of the secondary variants within the protein sequences compared to previously reported primary mutations within these genes (Figure 5C). In fact, in some cases, we observed stopgain secondary variants that were more premature in the protein sequence than previously reported disruptive mutations in these genes, suggesting that the second-hit can exert an effect as severe as the first-hit. The allelic diversity of second hits within these genes suggests that further functional analysis should be performed in order to understand their impact on phenotypic variability.

DISCUSSION

Our results support an oligogenic model, where the primary variant provides a sensitized genomic background for neurodevelopmental disease and secondary variants within functionally-intolerant genes determine the phenotypic trajectory of the disorder[6]. We propose that to surpass the genetic threshold for overt neurodevelopmental disease, molecular lesions need to accumulate, and that the number of second hits required for disease manifestation varies depending on the pathogenic effect of the first-hit. Some primary variants that are more tolerant to changes in the genetic background, such as the 16p12.1 deletion, are transmitted through generations and only surpass the threshold for severe disease with the accumulation of several rare pathogenic mutations[28,105]. We show that probands with 16p12.1 deletion have an average of two additional mutations affecting functionally intolerant genes compared to their carrier parents. Other variants such as 16p11.2 deletion, which are often *de novo*, push the genetic background closer to the threshold for severe manifestation and require lesser contribution from secondary variants. Similarly, rare syndromic CNVs such as Smith-Magenis syndrome and Sotos syndrome, which occur mostly *de novo* and encompass more functionally intolerant genes compared to variably expressive CNVs ($p=0.026$, one-tailed Mann-Whitney, Table S19, Figures S18), would push the genetic liability beyond the threshold for severe disease[6,36,106]. While additional mutations may not be necessary for complete penetrance of the disease, when present, these variants could modify the syndromic phenotypes. For instance, deleterious mutations in histone modifier genes have been reported to contribute to the conotruncal heart manifestation of 22q11.2 deletion syndrome[107]. This would also be the case for single-gene disorders where second hits potentially explain the discordant clinical features reported among affected carriers

of the same molecular alteration, as described for Rett syndrome and individuals with disruptive mutations in the intellectual disability gene *PACS1*[108,109].

Our results exemplify how variants in the genetic background not only contribute to a severe manifestation of the first-hit, but also to their variable clinical manifestation. This burden is even higher among probands with strong family history of neurodevelopmental or psychiatric disease, explaining their more severe clinical manifestation compared to probands with mild or negative family history of disease. These results provide an interpretation for the previously reported role of parental profiles towards the final clinical outcome in probands carrying inherited rare CNVs[89,90], and highlight the importance of interrogating the family history of psychiatric and neurodevelopmental disease for more accurate diagnostic assessment of the affected children.

We also found that secondary variants could modulate the manifestation of specific phenotypes. For example, in the presence of disease-associated rare CNVs or disruptive mutations in individual genes, secondary variants modulate FSIQ scores, where a high burden of second hits is likely to lead to severe cognitive deficits. However, the extent to which secondary variants can modulate a trait appears to depend upon the pathogenicity of the first-hit in that specific phenotypic domain. For example, we show that individuals with intellectual disability who carry 16p11.2 deletion have a disruption of only one additional functionally intolerant gene on average, likely enough to surpass the genetic threshold for intellectual disability. Moreover, the correlation observed between HC z-scores and burden of secondary variants could explain the incomplete penetrance of macrocephaly among carriers of the deletion [94,95,110,111].

Overall, these results suggest a multi-dimensional effect of secondary variants towards clinical features, and that their contribution to specific phenotypic domains depends on the extent

to which the primary variant sensitizes an individual towards a specific phenotypic trajectory. An important observation from our study is that a large number of disease-associated variants, that were deemed to be solely causative for the disorder, are in fact accompanied by substantial amount of rare genetic variation. Longitudinal and quantitative phenotyping across multiple developmental domains in all family members, along with whole genome sequencing studies in affected and asymptomatic individuals with a primary variant, are necessary for a more accurate understanding of these complex disorders. Therefore, in this new era of personalized medicine, it is critical that even after identifying a likely diagnostic disruptive mutation, further analysis of secondary variants are performed in order to provide appropriate counseling and management.

DESCRIPTION OF SUPPLEMENTAL DATA

Supplemental data includes 18 figures and 19 tables.

ACKNOWLEDGMENTS

We are grateful to all of the families in each cohort (16p12.1 deletion, SVIP and SSC) who participated in the study. This work was supported by NIH R01-GM121907, SFARI Pilot Grant (#399894) and resources from the Huck Institutes of the Life Sciences to S.G. We appreciate obtaining access to genomic and phenotypic data on SFARI Base. Approved researchers can obtain the SSC and SVIP population datasets described in this study by applying at https://base.sfari.org. L.P. was supported by Fulbright Commission Uruguay – ANII and the Huck Institutes of the Life Sciences. M.J. was supported by NIH T32-GM102057. C.R., L.C., O.G. and E.A. were supported by the Italian Ministry of Health and ‘5 per mille’ funding. Dedicated to the memory of Ethan Francis Schwartz, 1996-1998. K.M. is a Jacobs Foundation

Research Fellow. A.R. is supported by the Swiss National Science Foundation (31003A_160203).

WEB RESOURCES

OMIM, https://omim.org

Minitab Software, https://www.minitab.com.

JMP Pro 13.1.0 software, https://www.jmp.com

Sickle v.1.33, https://github.com/najoshi/sickle

SVIP, https://base.sfari.org/

REFERENCES

1. Vorstman JAS, Parr JR, Moreno-De-Luca D, Anney RJL, Nurnberger JI, Jr., Hallmayer JF. Autism genetics: opportunities and challenges for clinical translation. Nature reviews Genetics 2017;18:362-76.
2. Iossifov I, Ronemus M, Levy D, et al. De novo gene disruptions in children on the autistic spectrum. Neuron 2012;74:285-99.
3. Neale BM, Kou Y, Liu L, et al. Patterns and rates of exonic de novo mutations in autism spectrum disorders. Nature 2012;485:242-5.
4. O'Roak BJ, Vives L, Girirajan S, et al. Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. Nature 2012;485:246-50.
5. Sanders SJ, Murtha MT, Gupta AR, et al. De novo mutations revealed by whole-exome sequencing are strongly associated with autism. Nature 2012;485:237-41.
6. Girirajan S, Rosenfeld JA, Coe BP, et al. Phenotypic heterogeneity of genomic disorders and rare copy-number variants. New England Journal of Medicine 2012;367:1321-31.
7. Kendall KM, Rees E, Escott-Price V, et al. Cognitive Performance Among Carriers of Pathogenic Copy Number Variants: Analysis of 152,000 UK Biobank Subjects. Biol Psychiatry 2017;82:103-10.
8. Moreno-De-Luca A, Myers SM, Challman TD, Moreno-De-Luca D, Evans DW, Ledbetter DH. Developmental brain dysfunction: revival and expansion of old concepts based on new genetic evidence. The Lancet Neurology 2013;12:406-14.
9. Mannik K, Magi R, Mace A, et al. Copy number variations and cognitive phenotypes in unselected populations. Jama 2015;313:2044-54.
10. Zufferey F, Sherr EH, Beckmann ND, et al. A 600 kb deletion syndrome at 16p11. 2 leads to energy imbalance and neuropsychiatric disorders. Journal of medical genetics 2012;49:660-8.
11. Hanson E, Bernier R, Porche K, et al. The cognitive and behavioral phenotype of the 16p11.2 deletion in a clinically ascertained population. Biol Psychiatry 2015;77:785-93.
12. Bijlsma E, Gijsbers A, Schuurs-Hoeijmakers J, et al. Extending the phenotype of recurrent rearrangements of 16p11. 2: deletions in mentally retarded patients without autism and in normal individuals. European journal of medical genetics 2009;52:77-87.
13. Kumar RA, KaraMohamed S, Sudi J, et al. Recurrent 16p11. 2 microdeletions in autism. Human molecular genetics 2007;17:628-38.
14. Marshall CR, Noor A, Vincent JB, et al. Structural variation of chromosomes in autism spectrum disorder. The American Journal of Human Genetics 2008;82:477-88.
15. Walters R, Jacquemont S, Valsesia A, et al. A new highly penetrant form of obesity due to deletions on chromosome 16p11. 2. Nature 2010;463:671-5.
16. Weiss LA, Shen Y, Korn JM, et al. Association between microdeletion and microduplication at 16p11. 2 and autism. New England Journal of Medicine 2008;358:667-75.
17. 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. Journal of neurodevelopmental disorders 2010;2:26-38.
18. 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;19:1579-85.
19. D'Angelo D, Lebon S, Chen Q, et al. Defining the Effect of the 16p11.2 Duplication on Cognition, Behavior, and Medical Comorbidities. JAMA Psychiatry 2016;73:20-30.
20. Wu N, Ming X, Xiao J, et al. TBX6 null variants and a common hypomorphic allele in congenital scoliosis. The New England journal of medicine 2015;372:341-50.
21. Buxbaum JD, Cai G, Chaste P, et al. Mutation screening of the PTEN gene in patients with autism spectrum disorders and macrocephaly. American Journal of Medical Genetics Part B: Neuropsychiatric Genetics 2007;144:484-91.

22. Ching MS, Shen Y, Tan WH, et al. Deletions of NRXN1 (Neurexin□1) predispose to a wide spectrum of developmental disorders. American Journal of Medical Genetics Part B: Neuropsychiatric Genetics 2010;153:937-47.
23. Herman GE, Butter E, Enrile B, Pastore M, Prior TW, Sommer A. Increasing knowledge of PTEN germline mutations: Two additional patients with autism and macrocephaly. American journal of medical genetics Part A 2007;143:589-93.
24. Varga EA, Pastore M, Prior T, Herman GE, McBride KL. The prevalence of PTEN mutations in a clinical pediatric cohort with autism spectrum disorders, developmental delay, and macrocephaly. Genetics in Medicine 2009;11:111-7.
25. Butler MG, Dasouki M, Zhou X, et al. Subset of individuals with autism spectrum disorders and extreme macrocephaly associated with germline PTEN tumour suppressor gene mutations. Journal of medical genetics 2005;42:318-21.
26. Guerrini R, Cellini E, Mei D, et al. Variable epilepsy phenotypes associated with a familial intragenic deletion of the SCN1A gene. Epilepsia 2010;51:2474-7.
27. Suls A, Velizarova R, Yordanova I, et al. Four generations of epilepsy caused by an inherited microdeletion of the SCN1A gene. Neurology 2010;75:72-6.
28. Hammer MF, Ishii A, Johnstone L, et al. Rare variants of small effect size in neuronal excitability genes influence clinical outcome in Japanese cases of SCN1A truncation-positive Dravet syndrome. PLoS One 2017;12:e0180485.
29. Leslie NR, Longy M. Inherited PTEN mutations and the prediction of phenotype. Seminars in cell & developmental biology; 2016: Elsevier. p. 30-8.
30. Spinelli L, Black FM, Berg JN, Eickholt BJ, Leslie NR. Functionally distinct groups of inherited PTEN mutations in autism and tumour syndromes. Journal of medical genetics 2015;52:128-34.
31. Todarello G, Feng N, Kolachana BS, et al. Incomplete penetrance of NRXN1 deletions in families with schizophrenia. Schizophrenia research 2014;155:1-7.
32. Zori RT, Marsh DJ, Graham GE, Marliss EB, Eng C. Germline PTEN mutation in a family with Cowden syndrome and Bannayan□Riley□Ruvalcaba syndrome. American Journal of Medical Genetics Part A 1998;80:399-402.
33. St Clair D, Blackwood D, Muir W, et al. Association within a family of a balanced autosomal translocation with major mental illness. Lancet 1990;336:13-6.
34. Blackwood DH, Fordyce A, Walker MT, St Clair DM, Porteous DJ, Muir WJ. Schizophrenia and affective disorders--cosegregation with a translocation at chromosome 1q42 that directly disrupts brain-expressed genes: clinical and P300 findings in a family. American journal of human genetics 2001;69:428-33.
35. Blackwood DH, Muir WJ. Clinical phenotypes associated with DISC1, a candidate gene for schizophrenia. Neurotoxicity research 2004;6:35-41.
36. Girirajan S, Rosenfeld JA, Cooper GM, et al. A recurrent 16p12. 1 microdeletion supports a two-hit model for severe developmental delay. Nature genetics 2010;42:203-9.
37. De Vries B, White S, Knight S, et al. Clinical studies on submicroscopic subtelomeric rearrangements: a checklist. Journal of medical genetics 2001;38:145-50.
38. 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. European journal of human genetics : EJHG 2011;19:1152-60.
39. Qiao Y, Mercier E, Dastan J, et al. Copy number variants (CNVs) analysis in a deeply phenotyped cohort of individuals with intellectual disability (ID). BMC medical genetics 2014;15:82.
40. Stessman HA, Xiong B, Coe BP, et al. Targeted sequencing identifies 91 neurodevelopmental-disorder risk genes with autism and developmental-disability biases. Nat Genet 2017;49:515-26.
41. Iossifov I, O'Roak BJ, Sanders SJ, et al. The contribution of de novo coding mutations to autism spectrum disorder. Nature 2014;515:216-21.
42. Krumm N, Turner TN, Baker C, et al. Excess of rare, inherited truncating mutations in autism. Nature genetics 2015;47:582-8.

43. DePristo MA, Banks E, Poplin R, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet 2011;43:491-8.
44. McKenna A, Hanna M, Banks E, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 2010;20:1297-303.
45. Van der Auwera GA, Carneiro MO, Hartl C, et al. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. Current protocols in bioinformatics 2013;43:11 0 1-33.
46. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics 2010;26:589-95.
47. Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. Nat Genet 2014;46:310-5.
48. Petrovski S, Wang Q, Heinzen EL, Allen AS, Goldstein DB. Genic intolerance to functional variation and the interpretation of personal genomes. PLoS Genet 2013;9:e1003709.
49. Schwarz JM, Rodelsperger C, Schuelke M, Seelow D. MutationTaster evaluates disease-causing potential of sequence alterations. Nature methods 2010;7:575-6.
50. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res 2010;38:e164.
51. Robinson JT, Thorvaldsdóttir H, Winckler W, et al. Integrative genomics viewer. Nature biotechnology 2011;29:24-6.
52. Farrell MS, Werge T, Sklar P, et al. Evaluating historical candidate genes for schizophrenia. Mol Psychiatry 2015;20:555-62.
53. Fromer M, Pocklington AJ, Kavanagh DH, et al. De novo mutations in schizophrenia implicate synaptic networks. Nature 2014;506:179-84.
54. Gilissen C, Hehir-Kwa JY, Thung DT, et al. Genome sequencing identifies major causes of severe intellectual disability. Nature 2014;511:344-7.
55. Girard SL, Gauthier J, Noreau A, et al. Increased exonic de novo mutation rate in individuals with schizophrenia. Nature genetics 2011;43:860-3.
56. Greenwood TA, Light GA, Swerdlow NR, Radant AD, Braff DL. Association analysis of 94 candidate genes and schizophrenia-related endophenotypes. PLoS One 2012;7:e29630.
57. Grozeva D, Carss K, Spasic-Boskovic O, et al. Targeted Next-Generation Sequencing Analysis of 1,000 Individuals with Intellectual Disability. Hum Mutat 2015;36:1197-204.
58. Iossifov I, Levy D, Allen J, et al. Low load for disruptive mutations in autism genes and their biased transmission. Proceedings of the National Academy of Sciences of the United States of America 2015;112:E5600-7.
59. Kaufman L, Ayub M, Vincent JB. The genetic basis of non-syndromic intellectual disability: a review. Journal of neurodevelopmental disorders 2010;2:182-209.
60. Kirov G, Pocklington AJ, Holmans P, et al. De novo CNV analysis implicates specific abnormalities of postsynaptic signalling complexes in the pathogenesis of schizophrenia. Mol Psychiatry 2012;17:142-53.
61. Levy D, Ronemus M, Yamrom B, et al. Rare de novo and transmitted copy-number variation in autistic spectrum disorders. Neuron 2011;70:886-97.
62. O'Roak BJ, Vives L, Girirajan S, et al. Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. Nature 2012;485:246-50.
63. Purcell SM, Moran JL, Fromer M, et al. A polygenic burden of rare disruptive mutations in schizophrenia. Nature 2014;506:185-90.
64. Rauch A, Wieczorek D, Graf E, et al. Range of genetic mutations associated with severe non-syndromic sporadic intellectual disability: an exome sequencing study. Lancet 2012;380:1674-82.
65. Ripke S, O'Dushlaine C, Chambert K, et al. Genome-wide association analysis identifies 13 new risk loci for schizophrenia. Nat Genet 2013;45:1150-9.
66. Sanders SJ, He X, Willsey AJ, et al. Insights into Autism Spectrum Disorder Genomic Architecture and Biology from 71 Risk Loci. Neuron 2015;87:1215-33.

67. Sanders SJ, Murtha MT, Gupta AR, et al. De novo mutations revealed by whole-exome sequencing are strongly associated with autism. Nature 2012;485:237-41.
68. Xu B, Ionita-Laza I, Roos JL, et al. De novo gene mutations highlight patterns of genetic and neural complexity in schizophrenia. Nat Genet 2012;44:1365-9.
69. Wang K, Li M, Hadley D, et al. PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. Genome Res 2007;17:1665-74.
70. Sanders SJ, He X, Willsey AJ, et al. Insights into autism spectrum disorder genomic architecture and biology from 71 risk loci. Neuron 2015;87:1215-33.
71. Gonzalez-Mantilla AJ, Moreno-De-Luca A, Ledbetter DH, Martin CL. A Cross-Disorder Method to Identify Novel Candidate Genes for Developmental Brain Disorders. JAMA Psychiatry 2016;73:275-83.
72. Consortium GT. The Genotype-Tissue Expression (GTEx) project. Nat Genet 2013;45:580-5.
73. Greene CS, Krishnan A, Wong AK, et al. Understanding multicellular function and disease with human tissue-specific networks. Nat Genet 2015;47:569-76.
74. Krishnan A, Zhang R, Yao V, et al. Genome-wide prediction and functional characterization of the genetic basis of autism spectrum disorder. Nat Neurosci 2016;19:1454-62.
75. Kramer A, Green J, Pollard J, Jr., Tugendreich S. Causal analysis approaches in Ingenuity Pathway Analysis. Bioinformatics 2014;30:523-30.
76. Mi H, Muruganujan A, Casagrande JT, Thomas PD. Large-scale gene function analysis with the PANTHER classification system. Nature protocols 2013;8:1551-66.
77. Merico D, Isserlin R, Stueker O, Emili A, Bader GD. Enrichment map: a network-based method for gene-set enrichment visualization and interpretation. PLoS One 2010;5:e13984.
78. Shannon P, Markiel A, Ozier O, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 2003;13:2498-504.
79. Jacquemont S, Coe BP, Hersch M, et al. A higher mutational burden in females supports a “female protective model” in neurodevelopmental disorders. The American Journal of Human Genetics 2014;94:415-25.
80. Polyak A, Rosenfeld JA, Girirajan S. An assessment of sex bias in neurodevelopmental disorders. Genome medicine 2015;7:94.
81. Rees E, Walters JT, Chambert KD, et al. CNV analysis in a large schizophrenia sample implicates deletions at 16p12. 1 and SLC1A1 and duplications at 1p36. 33 and CGNL1. Human molecular genetics 2014;23:1669-76.
82. Stefansson H, Meyer-Lindenberg A, Steinberg S, et al. CNVs conferring risk of autism or schizophrenia affect cognition in controls. Nature 2014;505:361-6.
83. Grozeva D, Carss K, Spasic-Boskovic O, et al. De novo loss-of-function mutations in SETD5, encoding a methyltransferase in a 3p25 microdeletion syndrome critical region, cause intellectual disability. The American Journal of Human Genetics 2014;94:618-24.
84. Mefford HC, Batshaw ML, Hoffman EP. Genomics, intellectual disability, and autism. New England Journal of Medicine 2012;366:733-43.
85. Fukai R, Saitsu H, Okamoto N, et al. De novo missense mutations in NALCN cause developmental and intellectual impairment with hypotonia. Journal of human genetics 2016;61:451-5.
86. Gonzalez-Mantilla AJ, Moreno-De-Luca A, Ledbetter DH, Martin CL. A cross-disorder method to identify novel candidate genes for developmental brain disorders. JAMA psychiatry 2016;73:275-83.
87. Flanigan KM, Dunn DM, von Niederhausern A, et al. DMD Trp3X nonsense mutation associated with a founder effect in North American families with mild Becker muscular dystrophy. Neuromuscular disorders : NMD 2009;19:743-8.
88. Gurvich OL, Maiti B, Weiss RB, Aggarwal G, Howard MT, Flanigan KM. DMD exon 1 truncating point mutations: amelioration of phenotype by alternative translation initiation in exon 6. Hum Mutat 2009;30:633-40.

89. Moreno-De-Luca A, Evans DW, Boomer KB, et al. The role of parental cognitive, behavioral, and motor profiles in clinical variability in individuals with chromosome 16p11.2 deletions. JAMA Psychiatry 2015;72:119-26.
90. Olszewski AK, Radoeva PD, Fremont W, Kates WR, Antshel KM. Is child intelligence associated with parent and sibling intelligence in individuals with developmental disorders? An investigation in youth with 22q11.2 deletion (velo-cardio-facial) syndrome. Research in developmental disabilities 2014;35:3582-90.
91. Duyzend MH, Nuttle X, Coe BP, et al. Maternal Modifiers and Parent-of-Origin Bias of the Autism-Associated 16p11.2 CNV. American journal of human genetics 2016;98:45-57.
92. Simons Vip C. Simons Variation in Individuals Project (Simons VIP): a genetics-first approach to studying autism spectrum and related neurodevelopmental disorders. Neuron 2012;73:1063-7.
93. Bochukova EG, Huang N, Keogh J, et al. Large, rare chromosomal deletions associated with severe early-onset obesity. Nature 2010;463:666-70.
94. Jacquemont S, Reymond A, Zufferey F, et al. Mirror extreme BMI phenotypes associated with gene dosage at the chromosome 16p11.2 locus. Nature 2011;478:97-102.
95. Steinman KJ, Spence SJ, Ramocki MB, et al. 16p11.2 deletion and duplication: Characterizing neurologic phenotypes in a large clinically ascertained cohort. American journal of medical genetics Part A 2016;170:2943-55.
96. Desachy G, Croen L, Torres A, et al. Increased female autosomal burden of rare copy number variants in human populations and in autism families. Molecular psychiatry 2015;20:170-5.
97. Woodbury-Smith M, Nicolson R, Zarrei M, et al. Variable phenotype expression in a family segregating microdeletions of the NRXN1 and MBD5 autism spectrum disorder susceptibility genes. NPJ Genomic Medicine 2017;2:1.
98. Çelebi JT, Tsou HC, Chen FF, et al. Phenotypic findings of Cowden syndrome and Bannayan-Zonana syndrome in a family associated with a single germline mutation in PTEN. Journal of medical genetics 1999;36:360-4.
99. Bicknell LS, Morgan T, Bonafe L, et al. Mutations in FLNB cause boomerang dysplasia. J Med Genet 2005;42:e43.
100. Farrington-Rock C, Firestein MH, Bicknell LS, et al. Mutations in two regions of FLNB result in atelosteogenesis I and III. Hum Mutat 2006;27:705-10.
101. Mizutari K, Mutai H, Namba K, et al. High prevalence of CDH23 mutations in patients with congenital high-frequency sporadic or recessively inherited hearing loss. Orphanet J Rare Dis 2015;10:60.
102. Wang X, Li N, Xiong N, et al. Genetic Variants of Microtubule Actin Cross-linking Factor 1 (MACF1) Confer Risk for Parkinson's Disease. Molecular neurobiology 2017;54:2878-88.
103. Zhang D, Herring JA, Swaney SS, et al. Mutations responsible for Larsen syndrome cluster in the FLNB protein. J Med Genet 2006;43:e24.
104. Zhao C, Ikeda S, Arai T, et al. Association of the RYR3 gene polymorphisms with atherosclerosis in elderly Japanese population. BMC cardiovascular disorders 2014;14:6.
105. Moens LN, De Rijk P, Reumers J, et al. Sequencing of DISC1 pathway genes reveals increased burden of rare missense variants in schizophrenia patients from a northern Swedish population. PLoS One 2011;6:e23450.
106. Girirajan S, Eichler EE. Phenotypic variability and genetic susceptibility to genomic disorders. Human molecular genetics 2010;19:R176-R87.
107. Guo T, Chung JH, Wang T, et al. Histone Modifier Genes Alter Conotruncal Heart Phenotypes in 22q11.2 Deletion Syndrome. American journal of human genetics 2015;97:869-77.
108. Grillo E, Rizzo CL, Bianciardi L, et al. Revealing the complexity of a monogenic disease: Rett syndrome exome sequencing. PLoS One 2013;8:e56599.
109. Schuurs-Hoeijmakers JH, Oh EC, Vissers LE, et al. Recurrent de novo mutations in PACS1 cause defective cranial-neural-crest migration and define a recognizable intellectual-disability syndrome. American journal of human genetics 2012;91:1122-7.

110. Shinawi M, Liu P, Kang SH, et al. Recurrent reciprocal 16p11.2 rearrangements associated with global developmental delay, behavioural problems, dysmorphism, epilepsy, and abnormal head size. J Med Genet 2010;47:332-41.
111. Schaaf CP, Goin-Kochel RP, Nowell KP, et al. Expanding the clinical spectrum of the 16p11.2 chromosomal rearrangements: three patients with syringomyelia. European journal of human genetics : EJHG 2011;19:152-6.

FIGURES TITLES AND LEGENDS

Figure 1. Strategy for understanding the role of the genetic background in phenotypic variability of neurodevelopmental disease. A. Schematic of primary and secondary variants used in this study. Disease-associated mutations common among different individuals were considered as "primary variants" and likely pathogenic SNVs (colored X) or CNVs (colored boxes) affecting functionally intolerant genes were defined as "secondary variants". **B.** Combined clinical and genomic analysis of 757 probands and 233 family members carrying primary variants associated with disease (16p12.1 deletion, 16p11.2 deletion, 16 rare CNVs, *de novo* mutations in autism simplex cases, and inherited mutations in disease-associated genes) was performed to understand the role of rare (<0.1%) likely pathogenic secondary variants (SNVs with CADD≥25 and CNVs) in functionally intolerant genes (RVIS≤20th percentile) in the variable manifestation of neurodevelopmental disease.

Figure 2. Rare pathogenic mutations in functionally intolerant genes contribute to the phenotypic heterogeneity in 16p12.1 deletion. A. Phenotypic spectrum of 16p12.1 deletion in probands (n=141, red) and carrier parents (n=39, green). Probands exhibit a spectrum of severe developmental features compared to the mild cognitive and psychiatric features observed in carrier parents. Features represented were observed in ≥5% of probands or carrier parents. **B.** No overt common physical features were identified among carriers, suggesting the absence of facial features caused by the 16p12.1 microdeletion. Photographs of matched probands and 16p12.1 deletion carrier parents are shown. **C.** Private disruptive mutations in genes associated with neurodevelopmental disease, mostly occurring *de novo* or transmitted from non-carrier parents, ultimately lead to overt neurodevelopmental disease features in children with the deletion. DD: Developmental Delay; ASD: Autism Spectrum Disorder; ADHD: Attention Deficit

Hyperactivity Disorder; OCD: Obsessive-compulsive-like behavior. **D.** Global analysis of rare (≤0.1%) likely pathogenic secondary variants (SNVs with CADD≥25 or CNVs ≥50kbp) in genes intolerant to functional variation (RVIS≤20th percentile) in proband-carrier parent pairs with available exome and SNP array data. Probands present a higher second-hit mutation burden compared to their carrier parents (n=18, paired t-test, p=0.014).

Figure 3. Strong family history of neurodevelopmental and psychiatric disease is associated with excess of secondary variants and severe phenotypic outcome in 16p12.1 deletion probands. A. Diagram showing phenotypic heterogeneity in 56 probands with 16p12.1 deletion (black= phenotype present, white=absent, grey=not determined) and their family history of neurodevelopmental and psychiatric disease (red=strong family history, blue=mild/negative family history). **B.** Probands with strong family history of psychiatric and neurodevelopmental disease (n=9) have a more heterogeneous clinical manifestation (higher de Vries scores) than those with mild or negative family history (n=7) (one-tailed, Mann-Whitney, p=0.035). **C.** A higher burden of secondary variants (one-tailed Mann-Whitney p=0.001) and **D.** a greater difference in second-hit burden compared to carrier parents (p=0.0003) is found for probands with a strong family history (n=9) compared to those with a mild/negative family history (n=7). **E.** A higher proportion of secondary variants are transmitted from non-carrier parents in probands with strong family history compared to those with mild/negative family history (one-tailed Mann-Whitney, p=0.03).

Figure 4. Burden of secondary variants modulates quantitative phenotypes among probands with a first-hit CNV or SNV associated with neurodevelopmental disease. A. Negative correlation between the number of secondary variants and full-scale IQ (FSIQ) scores in individuals (n=53) carrying 16 CNVs associated with neurodevelopmental disease (Pearson

correlation, R=-0.36, p=0.004). Probands with 16p11.2 deletion (green), 16p11.2 duplication (red) 1q21.1 duplication (blue) and 7q11.23 duplication (yellow) are highlighted, while grey circles represent probands with other rare CNVs. **B.** Higher burden of secondary variants among probands with 16p11.2 deletion and FSIQ<70 (n=17), compared to probands with FSIQ≥70 (n=65, one-tailed, Mann-Whitney, p=0.08). Dots in the boxplots represent data points lower than 10th or higher than 90th percentiles. **C.** Negative correlation between the number of secondary variants and head circumference z-scores (age≥12 months, n=80, Pearson correlation R=-0.26, p=0.009), showing that while the 16p11.2 deletion leads to macrocephaly, secondary variants modulate this phenotype in the presence of the deletion. **D.** Autism probands with *de novo* disruptive mutations and available FSIQ scores (n=290) show a moderate negative correlation (Pearson correlation coefficient, R=-0.22, p=0.0001) between the number of secondary variants and FSIQ scores. **E.** Enrichment in the number of secondary variants in female probands with *de novo* mutations (n=46) compared to male probands (n=245, one-tailed Mann-Whitney, p=0.02) Dots in the boxplots represent data points lower than the 10th or higher than 90th percentile **F.** Probands present an excess of secondary variants compared to their unaffected siblings (n=184 pairs) matched for the same inherited disruptive mutations (loss-of function or missense CADD≥25) in genes recurrently disrupted in neurodevelopmental disease (paired t-test, p=0.031).

Figure 5. Secondary variants affect core biological processes and disease-associated genes. Genes with secondary variants found in **A.** autism probands carrying *de novo* disruptive mutations (SSC) and **B.** probands with the 16p11.2 deletion (SVIP) are enriched in core biological processes (FDR<0.05 with Bonferroni correction). Recurrent clusters of enriched gene ontology terms for “developmental processes”, “cell signaling”, “cell adhesion” and “transport”

functions are present among second hits found in each cohort. The size of each circle represents the number of genes annotated for each GO term; red shading of each circle represents the FDR for enrichment of each GO term among second-hit genes in each cohort, with darker shades indicating a lower FDR. Line thickness represents the number of shared annotated genes between pairs of GO terms**. C.** Location of variants in the protein sequences of *RIMS1*, *DIP2A*, *KDM5B* and *ACOX2* as examples of recurrent genes with secondary variants (green arrows) in probands with first-hit pathogenic CNVs or *de novo* disruptive mutations and previously reported to carry *de novo* disrupting mutations in simplex autism cases (red arrows).

Figure 1

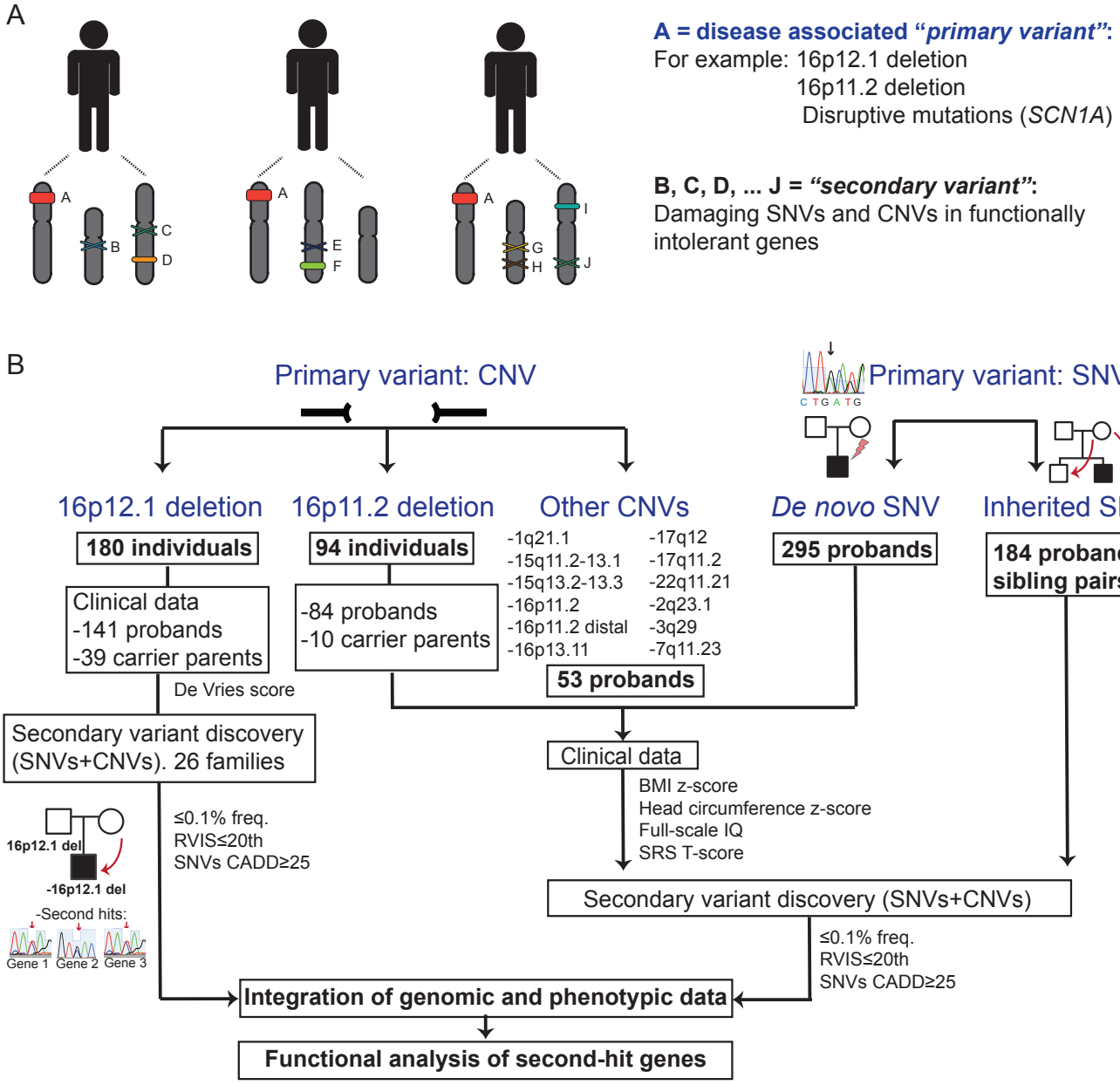
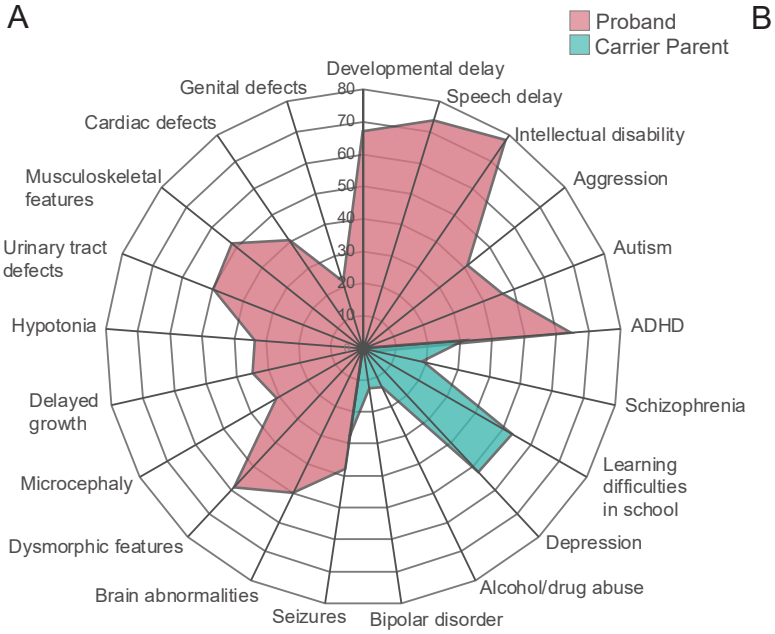
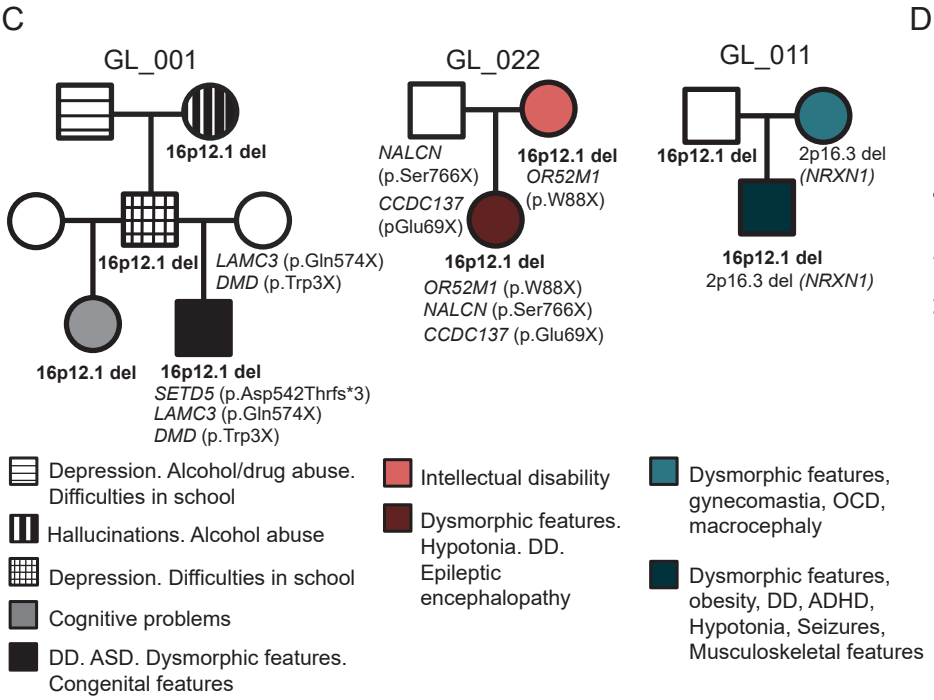

A
A
B
C
D
A
E
F
A
G
H
I
J
A = disease associated "primary variant":
For example: 16p12.1 deletion
16p11.2 deletion
Disruptive mutations (SCN1A)
B, C, D, ... J = "secondary variant":
Damaging SNVs and CNVs in functionally intolerant genes
B
Primary variant: CNV
Primary variant: SNV
C T G A T G
16p12.1 deletion
16p11.2 deletion
Other CNVs
De novo SNV
Inherited SNV
180 individuals
94 individuals
295 probands
184 proband/ sibling pairs
Clinical data
-141 probands
-39 carrier parents
-84 probands
-10 carrier parents
-1q21.1
-15q11.2-13.1
-15q13.2-13.3
-16p11.2
-16p11.2 distal
-16p13.11
-17q12
-17q11.2
-22q11.21
-2q23.1
-3q29
-7q11.23
53 probands
De Vries score
Secondary variant discovery (SNVs+CNVs). 26 families
Clinical data
BMI z-score
Head circumference z-score
Full-scale IQ
SRS T-score
≤0.1% freq.
RVIS≤20th
SNVs CADD≥25
16p12.1 del
-16p12.1 del
-Second hits:
Gene 1
Gene 2
Gene 3
Secondary variant discovery (SNVs+CNVs)
≤0.1% freq.
RVIS≤20th
SNVs CADD≥25
Integration of genomic and phenotypic data
Functional analysis of second-hit genes

Figure 2


A
B
Proband
Carrier Parent
Developmental delay
Speech delay
Intellectual disability
Aggression
Autism
ADHD
Schizophrenia
Learning difficulties in school
Depression
Alcohol/drug abuse
Bipolar disorder
Seizures
Brain abnormalities
Dysmorphic features
Microcephaly
Delayed growth
Hypotonia
Urinary tract defects
Musculoskeletal features
Cardiac defects
Genital defects
80
70
60
50
40
30
20
10


C
GL_001
16p12.1 del
16p12.1 del
LAMC3 (p.Gln574X)
DMD (p.Trp3X)
16p12.1 del
16p12.1 del
SETD5 (p.Asp542Thrfs*3)
LAMC3 (p.Gln574X)
DMD (p.Trp3X)
GL_022
NALCN (p.Ser766X)
CCDC137 (pGlu69X)
16p12.1 del
OR52M1 (p.W88X)
16p12.1 del
OR52M1 (p.W88X)
NALCN (p.Ser766X)
CCDC137 (p.Glu69X)
GL_011
16p12.1 del
2p16.3 del (NRXN1)
16p12.1 del
2p16.3 del (NRXN1)
Depression. Alcohol/drug abuse. Difficulties in school
Hallucinations. Alcohol abuse
Depression. Difficulties in school
Cognitive problems
DD. ASD. Dysmorphic features. Congenital features
Intellectual disability
Dysmorphic features. Hypotonia. DD. Epileptic encephalopathy
Dysmorphic features, gynecomastia, OCD, macrocephaly
Dysmorphic features, obesity, DD, ADHD, Hypotonia, Seizures, Musculoskeletal features

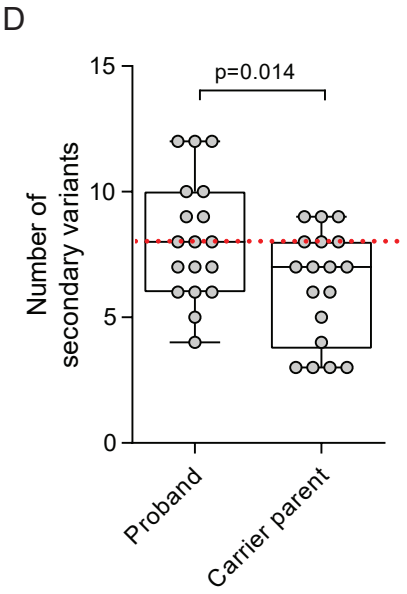

D
p=0.014
Number of secondary variants
15
10
5
0
Proband
Carrier parent

Figure 3

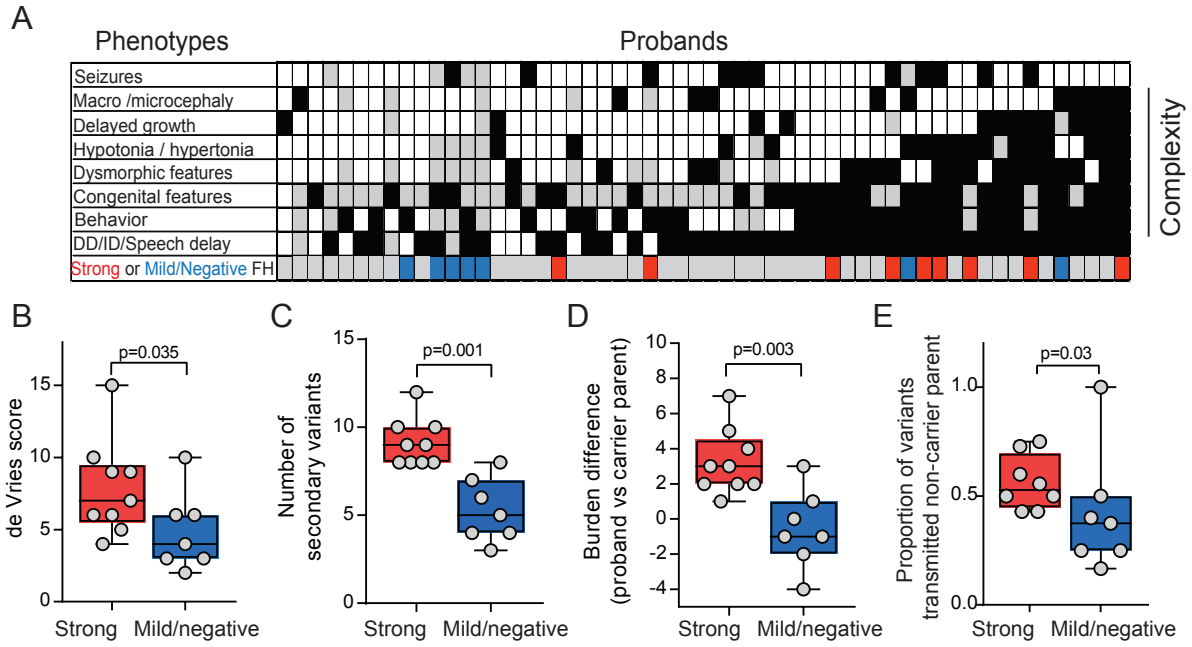
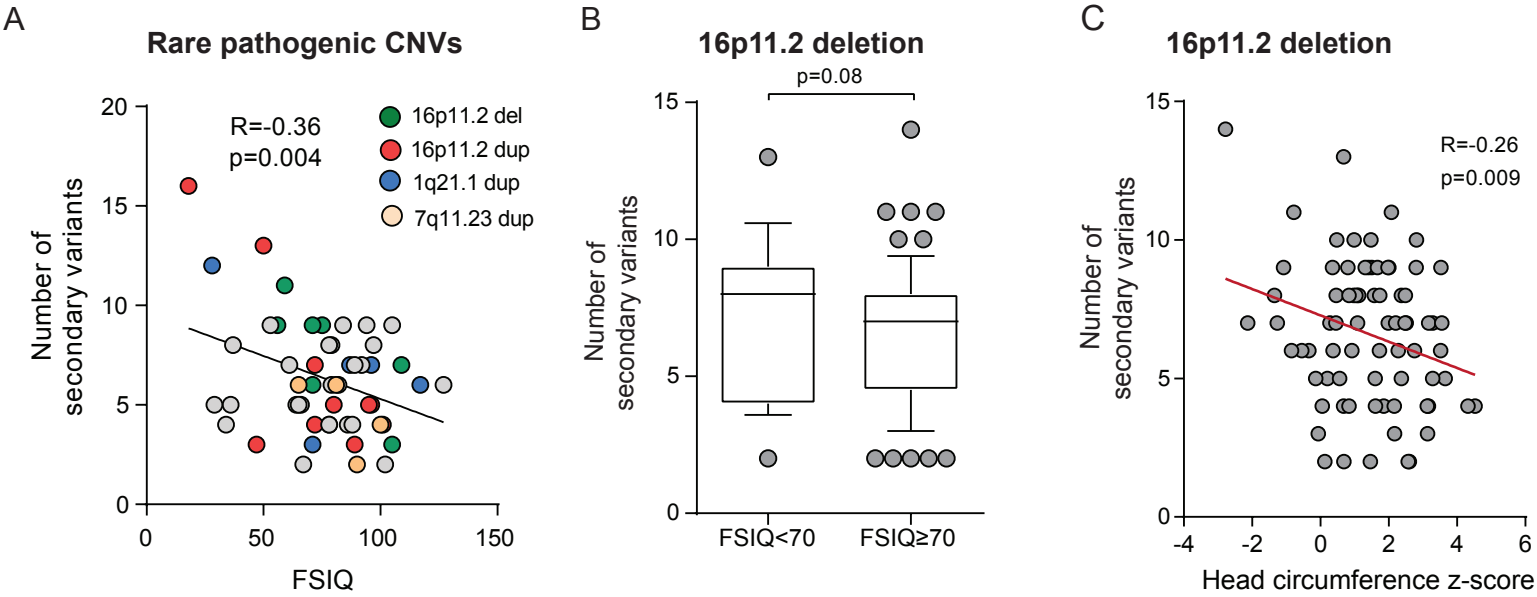

A
Phenotypes
Probands
Seizures
Macro /microcephaly
Delayed growth
Hypotonia / hypertonia
Dysmorphic features
Congenital features
Behavior
DD/ID/Speech delay
Strong or Mild/Negative FH
Complexity
B
de Vries score
p=0.035
Strong
Mild/negative
C
Number of secondary variants
p=0.001
Strong
Mild/negative
D
Burden difference (proband vs carrier parent)
p=0.003
Strong
Mild/negative
E
Proportion of variants transmitted non-carrier parent
p=0.03
Strong
Mild/negative

Figure 4

Primary variant CNV:



Primary variant SNV:

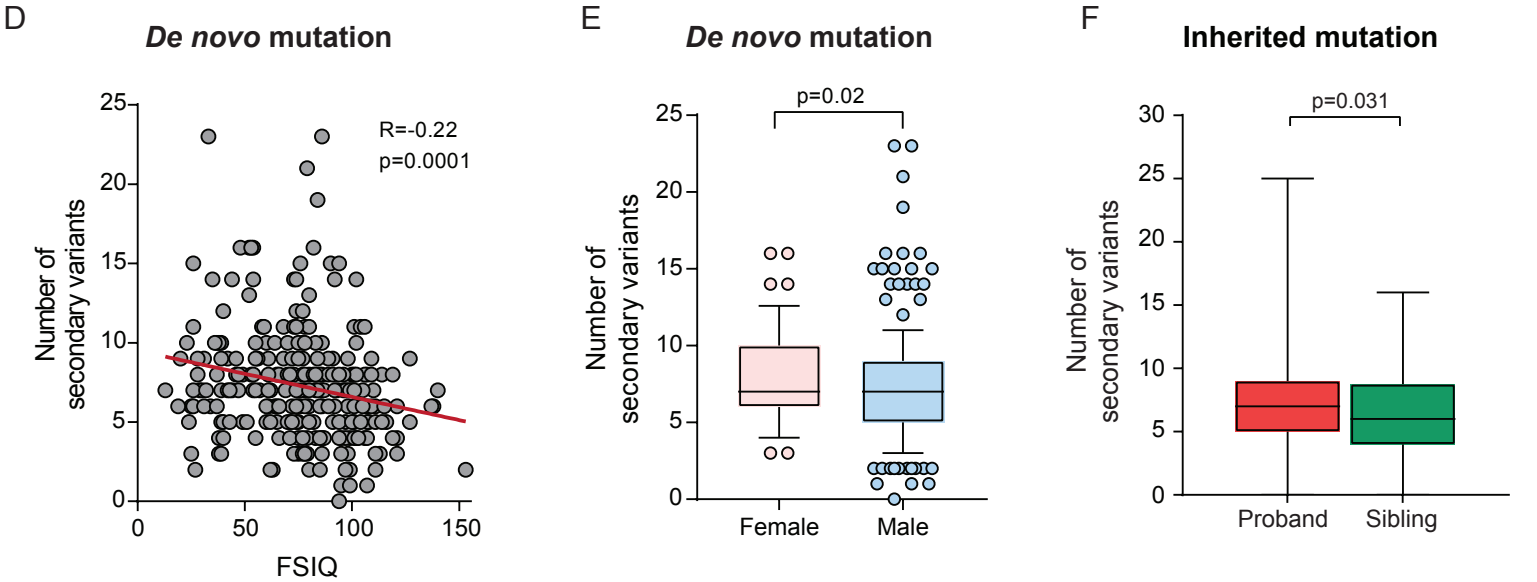
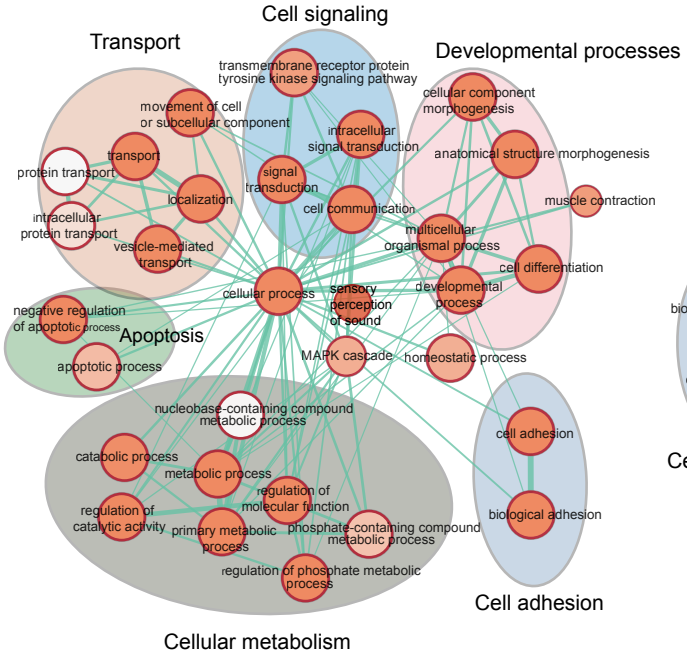


Figure 5

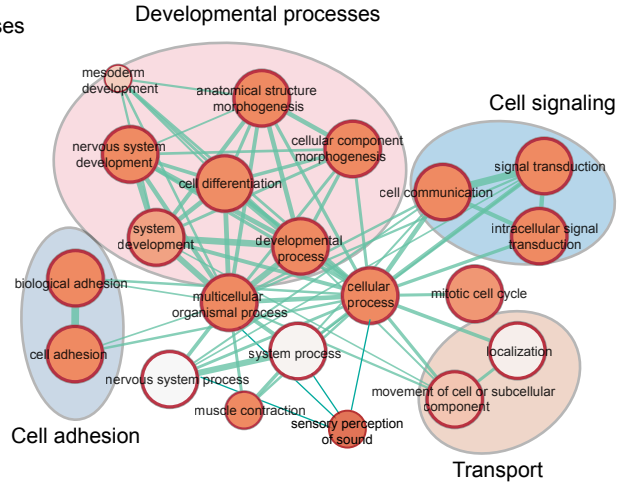
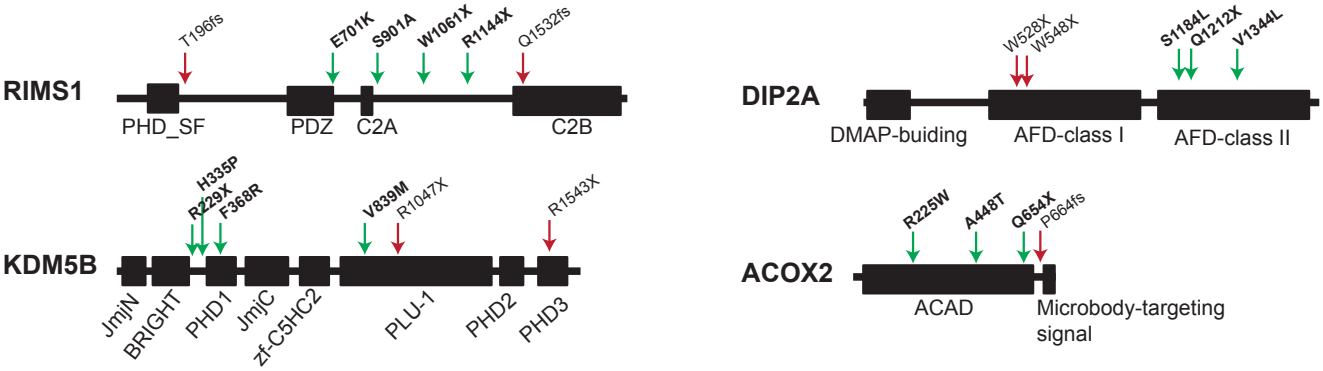
A

ASD probands with *de novo* mutations



B

16p11.2 deletion probands



C **Location of mutations in recurrent second-hits**