Rare variants in the genetic background modulate cognitive and developmental phenotypes in individuals carrying disease-associated variants

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,*]

Table of Contents

Supplementary Materials and Methods 5

1.1. Patient recruitment and clinical data ascertainment 5

1.2. Exome sequencing and SNP arrays 7

1.3. Functional analysis of rare variants in the genetic background 11

Supplementary Figures 13

Figure S1. 16p12.1 family pedigrees. 15

Figure S2. Enrichment of rare variants in the genetic background in probands compared to carrier parents and carrier siblings. 16

Figure S3. Other hits in 16p12.1 deletion probands are enriched for genes with high expression in the brain. 18

Figure S4. Family history of neurodevelopmental and psychiatric disease is associated with a higher burden of other hits and disease heterogeneity in probands with 16p12.1 deletion. 19

Figure S5. Other-hit burden correlates with FSIQ in SSC cohort individuals with rare CNVs associated with neurodevelopmental disease. 20

Figure S6. Number of other hits correlates with cognitive phenotypes in carriers of CNVs associated with neurodevelopmental phenotypes. 21

Figure S7. Rare variants in the genetic background and FSIQ scores in SVIP cohort probands with 16p11.2 deletion. 22

Figure S8. Phenotypic ascertainment differences among probands from the SSC and SVIP cohorts. 23

Figure S9. Number of other hits does not correlate with BMI or SRS T-scores in probands with rare CNVs (SSC cohort) or 16p11.2 deletion (SVIP cohort). 25

Figure S10. Number of rare variants in the genetic background in probands with *de novo* disruptive variants correlates with FSIQ and SRS T-scores. 26

Figure S11. Higher other-hit burden in female probands with *de novo* disruptive variants compared to male probands. 27

Figure S12. Known neurodevelopmental genes carrying other hits in probands with pathogenic CNVs or *de novo* disruptive variants. 28

Figure S13. Syndromic CNVs present lower RVIS scores compared to variably expressive CNVs. 29

Figure S14. Differences in FSIQ and number of other hits between probands with inherited and *de novo* 16p11.2 deletion. 30

Figure S15. Pipeline for the identification of other hits from exome sequencing and SNP array data. 31

Figure S16. No correlation between the length of genome with coverage ≥8X and the number of other-hit SNVs for individuals with 16p12.1 deletion. 32

Figure S17. Sample size calculation for 16p12.1 cohort. 33

Supplementary Tables 34

Table S1. Modified de Vries scoring rubric for uniform quantification of clinical heterogeneity and severity 34

Table S2. A summary of inheritance data for 16p12.1 deletion 35

Table S3. Phenotypes in probands with 16p12.1 deletion (Excel file) 35

Table S4. Phenotypes in carrier and non-carrier parents of probands with 16p12.1 deletion 36

Table S5. Summary of exome sequencing and SNP arrays performed on 16p12.1 probands and family members 38

Table S6. Coding variants in 16p12.1 genes identified on the non-deleted chromosome of probands with 16p12.1 deletion 39

Table S7. *De novo* variants identified in probands with 16p12.1 deletion and carrier siblings 40

Table S8. Private and rare gene-disruptive variants in disease-associated genes identified in probands with 16p12.1 deletion. 41

Table S9. Copy-number variants in families with 16p12.1 deletion (Excel file) 43

Table S10. Canonical pathway analysis of genes with identified other hits in probands and carrier parents with 16p12.1 deletion. 43

Table S11. Number of probands from the SSC cohort carrying rare CNVs associated with neurodevelopmental phenotypes 44

Table S12. Inherited disruptive variants in neurodevelopmental disease-associated genes found in autism proband-unaffected sibling pairs in the SSC cohort (Excel file) 44

Table S13. Genes originally identified to carry *de novo* disruptive variants in autism simplex cases but identified as other hits in this study (in probands with first-hit pathogenic CNVs or SNVs) 45

Table S14. Genes with other hits associated with skeletal, muscular, cardiovascular or renal disease (Excel file) 46

Table S15. Biological processes enriched among genes carrying other hits in SSC probands with *de novo* gene-disruptive variants 47

Table S16. Biological processes enriched among genes with other hits in SVIP probands with 16p11.2 deletion 48

Table S17. Syndromic and variably expressive CNVs associated with neurodevelopmental disorders 49

Table S18. Genetic variants found in individuals with 16p12.1 deletion and family members (Excel file)..... 50

Table S19. Other hits identified from exome-sequencing data in 16p12.1 deletion cohort (Excel file)..... 50

Supplementary References..... 51

Supplementary Materials and Methods

1.1. Patient recruitment and clinical data ascertainment

We analyzed clinical and genetic data in individuals carrying a disease-associated "primary variant" or "first-hit", defined as follows: (a) rare CNVs previously associated with neurodevelopmental disorders[1], (b) previously reported *de novo* pathogenic variants in candidate genes[2,3], and (c) inherited pathogenic variants in genes associated with neurodevelopmental disorders[4]. Families with 16p12.1 deletion (OMIM#136570) were recruited through medical genetics collaborators worldwide. Affected individuals from these families were diagnosed to carry a 16p12.1 deletion in a certified clinical diagnostic laboratory. The deletion was identified by array comparative genomic hybridization (aCGH) and then confirmed by fluorescence in situ hybridization (FISH) analysis. When available, all direct family members and extended family carrying the 16p12.1 deletion were recruited. Recruitment criteria for individuals with 16p12.1 deletion, individuals from simplex autism cases from the SSC cohort and individuals with 16p11.2 deletion from the SVIP cohort explicitly excluded individuals with other known genetic or Mendelian disorders. Proband/siblings pairs from the SSC cohort carrying pathogenic variants were identified as those carrying rare (≤0.1%) loss-of-function or likely damaging missense variants (Phred-like CADD≥25, representing the top 0.3% most deleterious variants in the genome)[5] in genes with recurrent *de novo* pathogenic variants from large cohorts of patients with neurodevelopmental disorders (≥2 *de novo* cases, classified as tier 1 and tier 2 by Gonzalez-Mantilla et al)[4].

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. 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 also 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.

Phenotypic severity and variability in 16p12.1 deletion probands was measured using a modified version of the de Vries scoring system. 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[6-8]. 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.

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 that 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.

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* pathogenic variants from the Simons Simplex Collection (SSC)[2,3]. FSIQ, SRS T-scores, BMI and head circumference z-scores were obtained from the Simons Foundation for 86 individuals from families with 16p11.2 BP4-BP5 deletion collected as part of the Simons Variation in Individuals Project (SVIP).

1.2. Exome sequencing and SNP arrays

We generated exome sequencing and SNP microarray data for 105 individuals from the first set of 26 families recruited with the 16p12.1 deletion using standardized pipelines[9-11]. 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)[12]. 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[9-11]. In order to detect splice-site variants 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 in 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 deleteriousness of the alternate allele (Mutation Taster, CADD score), allele frequency in the Exome Aggregation Consortium (ExAC database), and Residual Variation Intolerance Score (RVIS)[5,13-15]. 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) (Figure S15). 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 S18). Loss-of-function variants (LoF), including stopgain, frameshift insertion or deletion, and splice-site variants (predicted by MutationTaster as *disease causing*, "D", or *disease causing automatic*, "A"), as well as *de novo* variants in probands were visually

confirmed basis using Integrative Genomics Viewer (IGV)[16]. Rare (ExAC≤0.1%) missense (with Phred-like CADD ≥25) and LoF variants were investigated in sets of genes associated with neurodevelopmental disorders or reported as disease causing in OMIM[2,17-34]. A subset of disruptive variants 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[35]. 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[1] as determined by 50% reciprocal overlap), and overlap with segmental duplications and centromere/telomere sequences (≤75%) (Figure S15). CNV calls in children and parents were visually confirmed by inspection of log-R ratio (LRR) and B-allele frequency (BAF) plots.

Single-nucleotide variants and CNVs from autism simplex probands and unaffected siblings (from the SSC cohort) and 16p11.2 deletion probands (from the SVIP cohort) were filtered following the same procedures as for the 16p12.1 deletion cohort.

We restricted our search for other hits to a subset of genes less likely to harbor variants in a control population, as a proxy for association of genes with disease. The Residual Variation Intolerance Score (RVIS) has been shown to be a good predictor of gene intolerance to

deleterious variants and has been widely used by multiple studies for the recapitulation of known and the discovery of novel disease-associated genes[13]. For example, Krumm *et al*. showed that autism-associated genes have an average RVIS of 26th percentile, and using classifications from the human disease network, Petrovski *et al*. showed that genes involved in developmental, skeletal, cardiac, neurological and muscular disorders have average scores in the 20th-30th percentile[3,13]. Therefore, to focus our assessment of other hits towards a subset of genes relevant to disease pathogenicity, we defined "other hits" to only include rare likely deleterious variants in genes with RVIS≤20th percentile. The burden of other hits was measured for each individual as the number of functionally intolerant genes (with RVIS≤20th percentile) either carrying rare (frequency in ExAC database ≤0.1%) likely deleterious variants (loss-of-function and missense variants with a Phred-like CADD ≥25, representing the top 0.3% of most deleterious variants in the human genome) or within rare CNVs (found in ≤0.1% of a control population and ≥50 kbp)[5,13] (Table S19, Figure S17). RVIS scores (version 3) were downloaded from http://genic-intolerance.org. Following Exome Variant Server recommendations, RVIS percentiles were calculated from the combination of European and African populations, with MAF filter set at 0.1%. We did not observe any correlation between the number of other hits 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 other hits within each cohort (Figure S16). 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 v. 13.1.0.

Proband-parent or proband-sibling burden of other hits in the genetic background and clinical severity scores were compared using the Wilcoxon signed-rank test. Genetic burden and

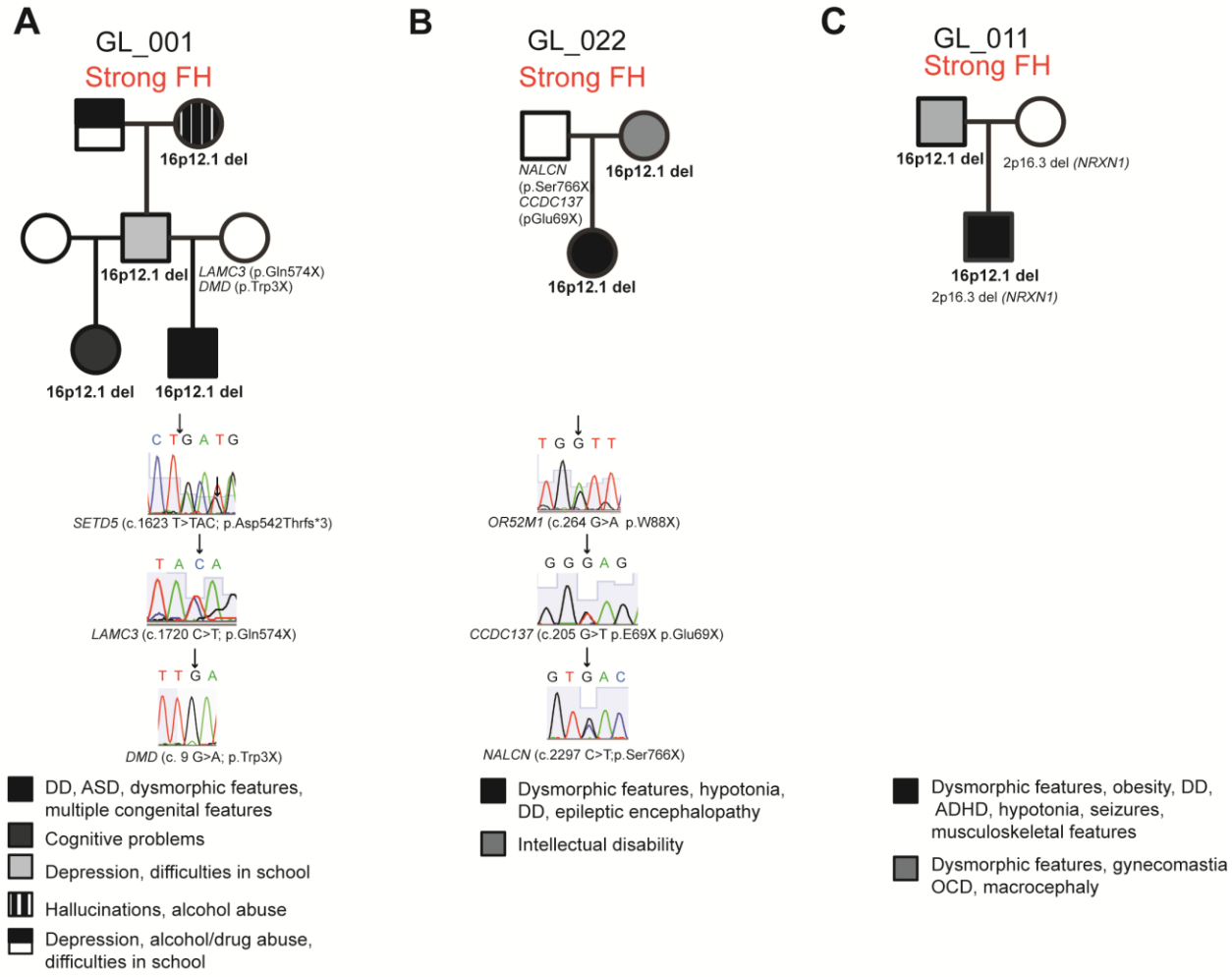
clinical severity scores between different categories of probands with related or shared primary variants were compared using non-parametric one-tailed Mann-Whitney tests, due to the hypothesis-driven nature of the comparison. Equal variances between groups were statistically compared using test for equal variances (ANOVA). The Kolmogorov-Smirnov test was used to assess the normality distribution of the burden of other hits, FSIQ, SRS T-scores, BMI z-scores and head circumference (HC) z-scores. Correlation between the number of other hits and quantitative phenotypes was assessed using Pearson's correlation for normally distributed datasets, or Spearman's correlation for datasets that were not normally distributed. Statistics were calculated using Minitab or R (v.3.4.2) software. Boxplots presented in the results display the distribution of data from minimum to maximum.

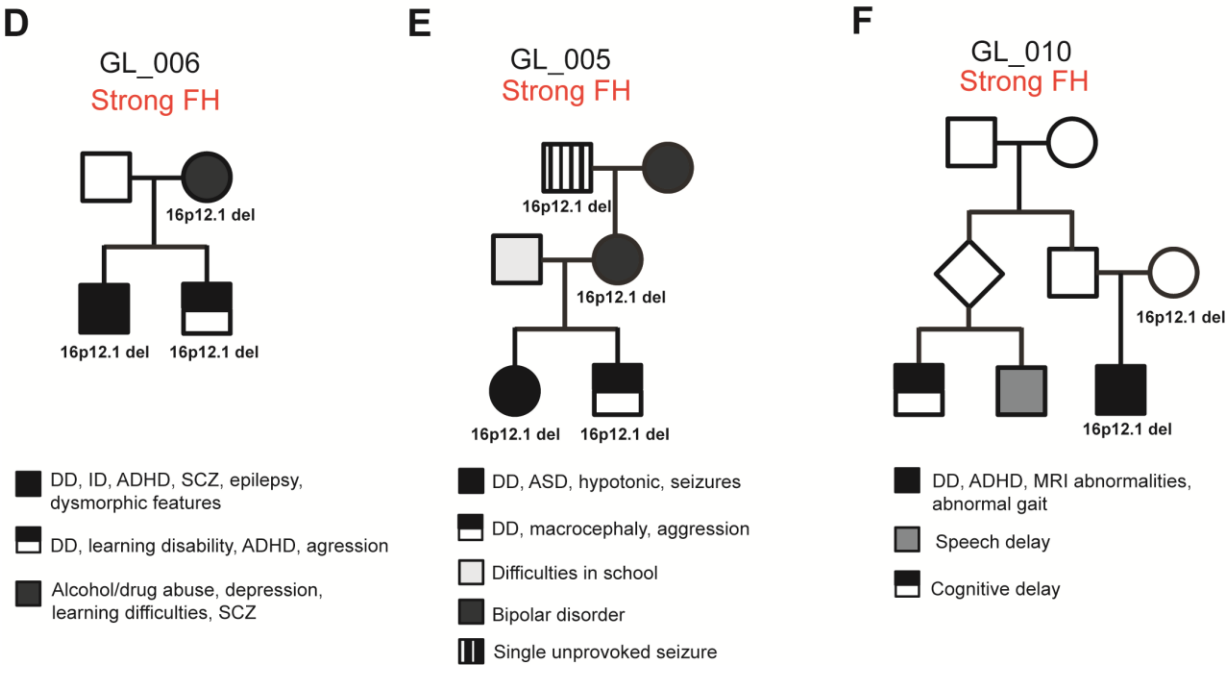
1.3. Functional analysis of rare variants in the genetic background

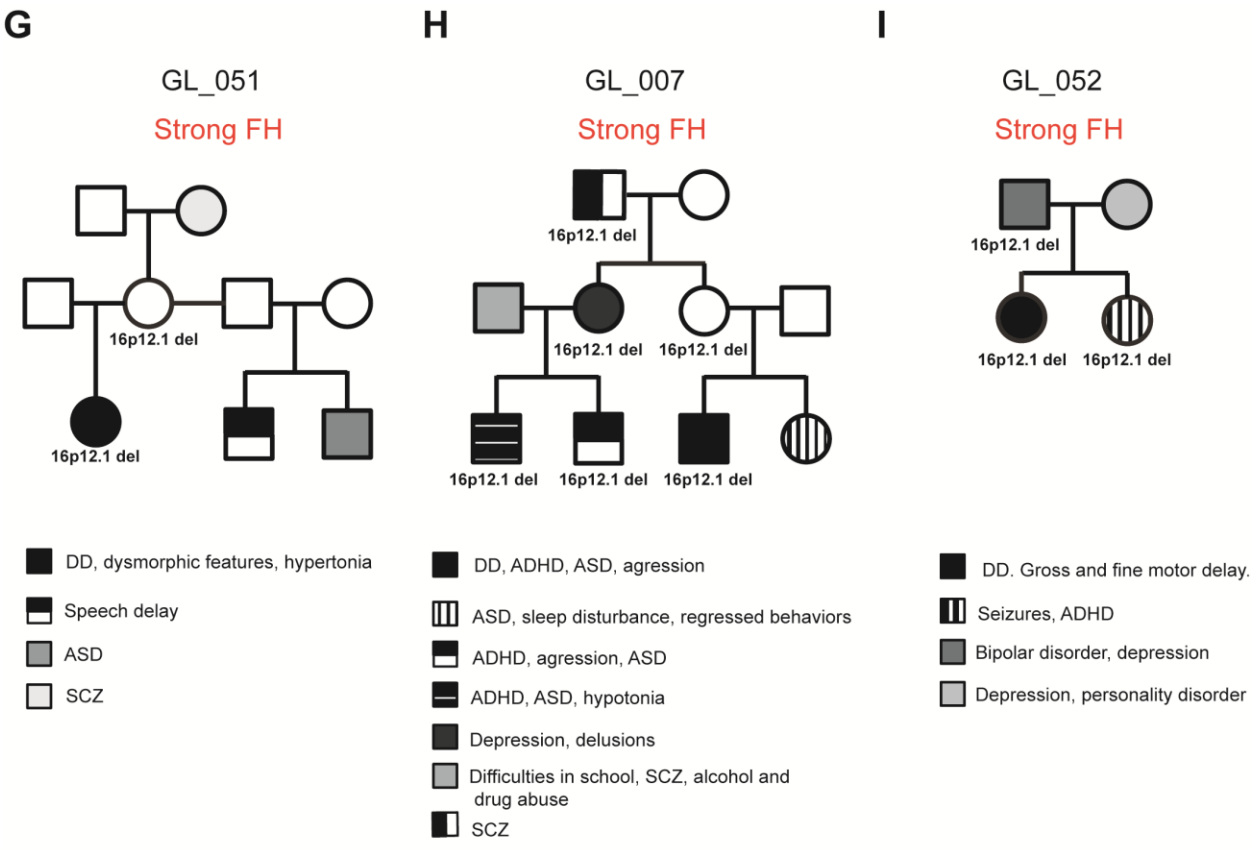
To identify enrichment in specific canonical pathways among the genes conferring a higher burden in probands with 16p12.1 deletion, we performed IPA on 219 genes carrying other variants in probands and 130 genes in carrier parents, using the Ingenuity Knowledge Base as a reference set (QIAGEN)[36] . 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. Tissue-specific median RPKM expression values for human genes were obtained from the GTex database[37]. 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 other hits in genes with high expression in brain or non-brain tissues (GTex) was

compared between probands and carrier parents using Wilcoxon signed-rank test. Gene ontology (GO) enrichment analysis of genes with other hits identified in SSC probands with *de novo* pathogenic variants and 16p11.2 BP4-BP5 deletion probands from SVIP was performed using the Panther Statistical Overrepresentation test[38]. 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[39,40]

Supplementary Figures






G
GL_051
Strong FH
16p12.1 del
16p12.1 del
DD, dysmorphic features, hypertonia
Speech delay
ASD
SCZ
H
GL_007
Strong FH
16p12.1 del
16p12.1 del
16p12.1 del
16p12.1 del
16p12.1 del
16p12.1 del
DD, ADHD, ASD, agression
ASD, sleep disturbance, regressed behaviors
ADHD, agression, ASD
ADHD, ASD, hypotonia
Depression, delusions
Difficulties in school, SCZ, alcohol and drug abuse
SCZ
I
GL_052
Strong FH
16p12.1 del
16p12.1 del
16p12.1 del
DD. Gross and fine motor delay.
Seizures, ADHD
Bipolar disorder, depression
Depression, personality disorder
J
GL_012
Mild FH
16p12.1 del
16p12.1 del
Obese, mild speech and motor delay, ADHD, pervasive developmental disorder, sleep disturbance, macrocephaly, Becker's muscular dystrophy
Single unprovoked seizure
Depression
K
GL_015
Negative FH
16p12.1 del
16p12.1 del
DD, intellectual disability, ADHD, agression, ASD, sleep disturbance
L
GL_046
Negative FH
16p12.1 del
16p12.1 del
Psychomotor delay, congenital cardiac malformation

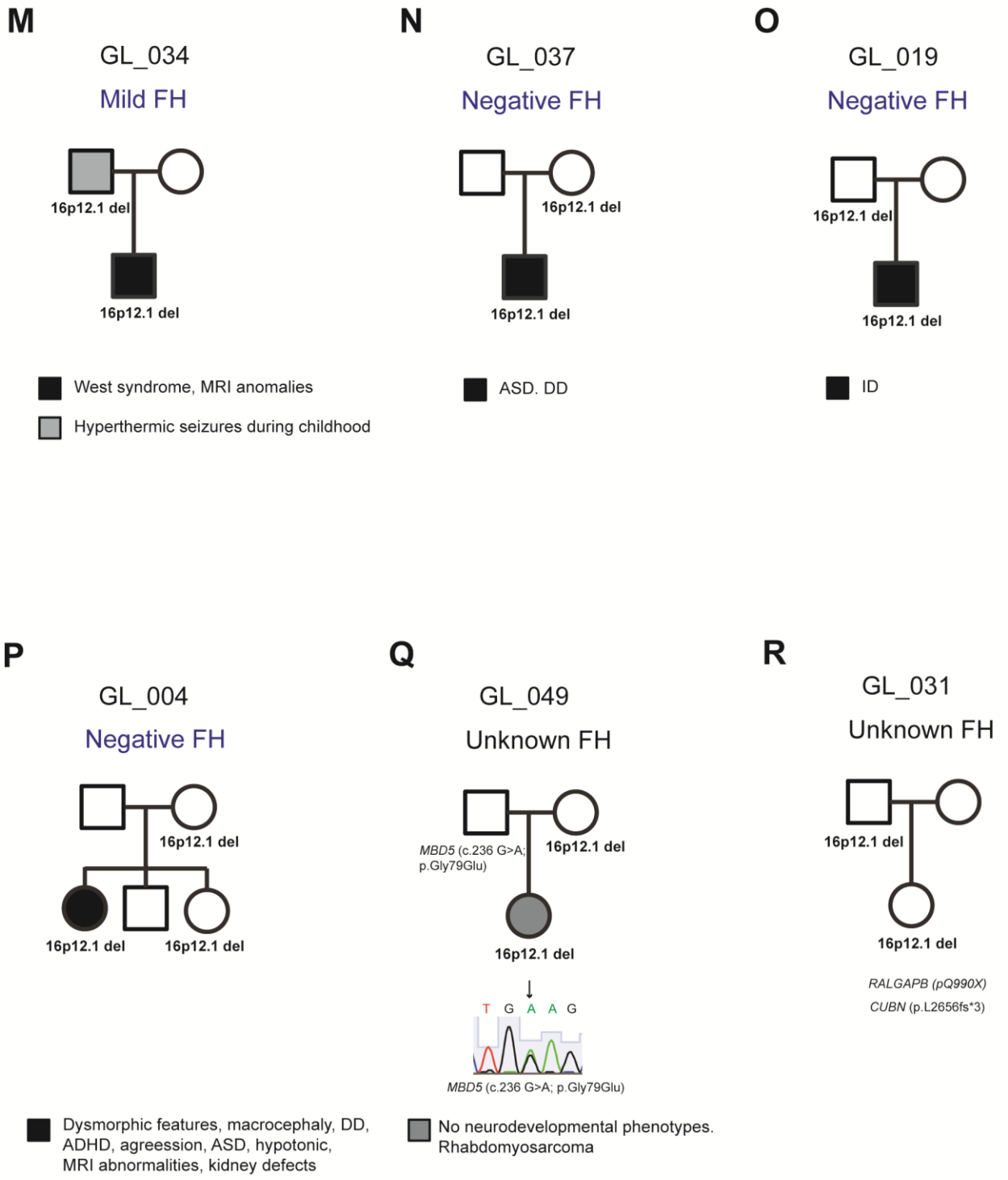


Figure S1. 16p12.1 family pedigrees. Pedigrees of families with 16p12.1 deletion with known family history of neuropsychiatric disease and/or validated likely deleterious variants in disease-associated genes. Variants identified are shown with Sanger validations. “DD”, developmental delay; “ADHD”, attention deficit hyperactivity disorder; “ASD”, autism spectrum disorder; “SCZ”, schizophrenia.

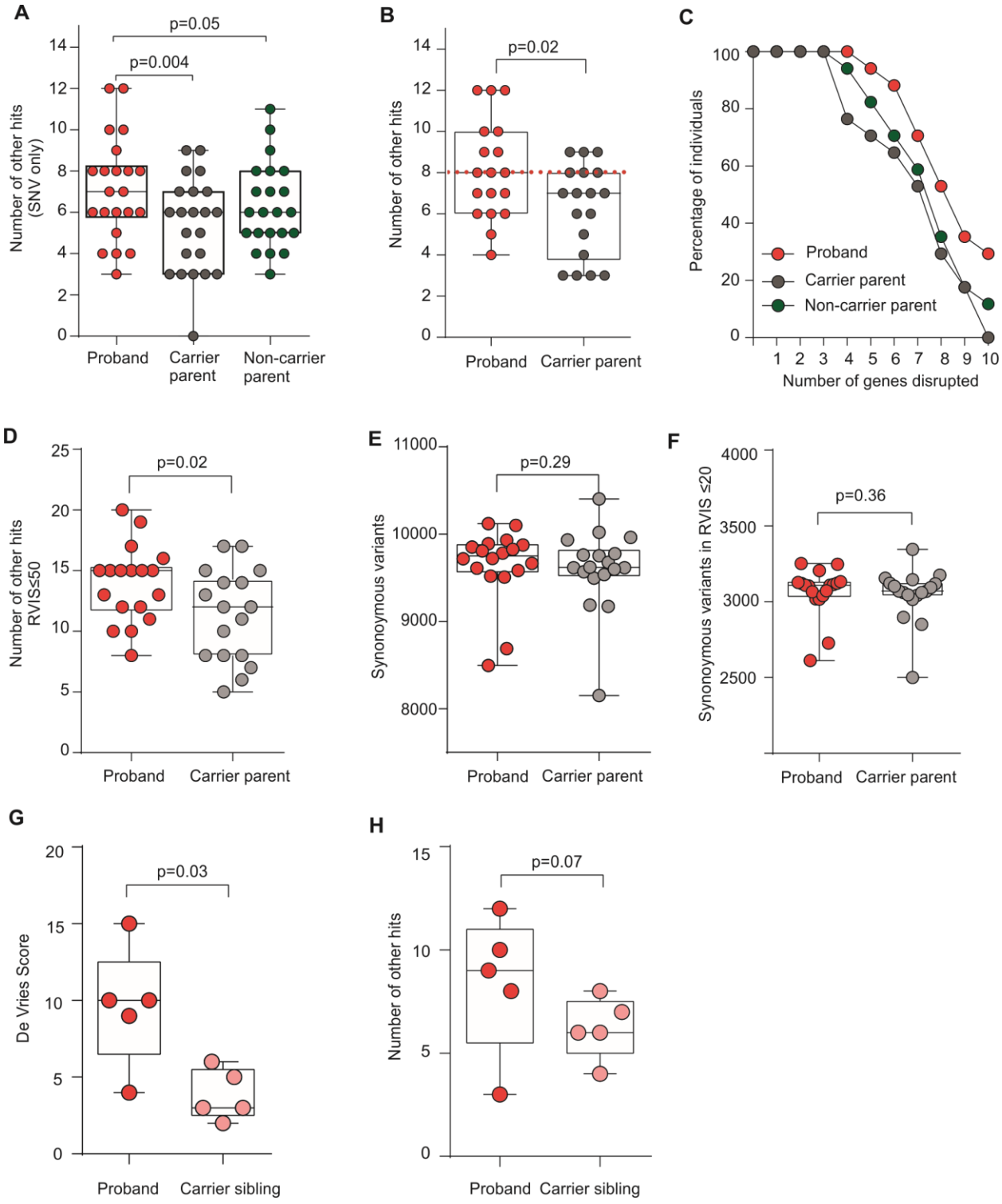


Figure S2. Enrichment of rare variants in the genetic background in probands compared to carrier parents and carrier siblings. A. Excess of other hits (SNV only) in probands with 16p12.1 deletion compared to their carrier parents (n=22, Wilcoxon signed-rank test, p=0.004), with a marginal difference compared to their non-carrier parents (p=0.05), suggests a higher contribution of other hits in probands from non-carrier parents. **B.** Higher burden of other hits

(SNVs with CADD≥25 and CNVs) in probands compared to their carrier parents (n=18, Wilcoxon signed-rank test, p=0.02). **C.** A higher percentage of probands with 16p12.1 deletion carry any number of rare (≤0.1%) likely deleterious variants (CNVs or SNVs) affecting functionally intolerant genes (RVIS≤20) compared to carrier parents (n=18 pairs), which becomes significant at 10 variants (one-tailed Fisher's exact test, p=0.02). **D.** A significantly higher burden of other hits is observed in probands when functional intolerance is defined at RVIS ≤50th percentile (n=18, Wilcoxon signed-rank test, p=0.02). **E.** No change in the overall number of synonymous variants (Wilcoxon signed-rank test, p=0.29) or **F.** synonymous variants affecting functionally intolerant genes (Wilcoxon signed-rank test, p=0.36) in probands compared to carrier parents. **G.** Probands present higher de Vries scores compared to carrier siblings with 16p12.1 deletion (p=0.03, Wilcoxon singed-rank test). **H.** Probands present a non-significant increase in the number of other hits compared to their mildly affected carrier siblings (p=0.07, Wilcoxon singed-rank test).

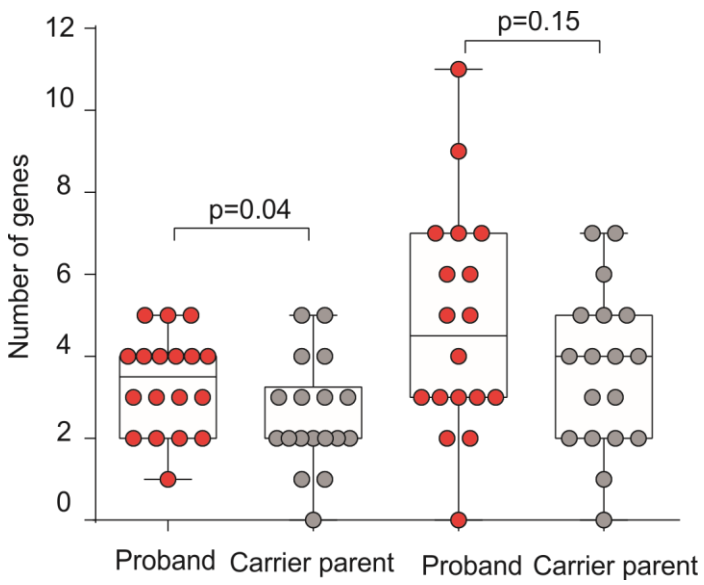


Figure S3. Other hits in 16p12.1 deletion probands are enriched for genes with high expression in the brain. Probands present an excess of other hits in genes with preferential expression in the brain compared to their carrier parents (n=18, Wilcoxon signed-rank test p=0.04), while no difference was observed for genes not preferentially expressed in the brain (n=18, Wilcoxon signed-rank test, p=0.15).

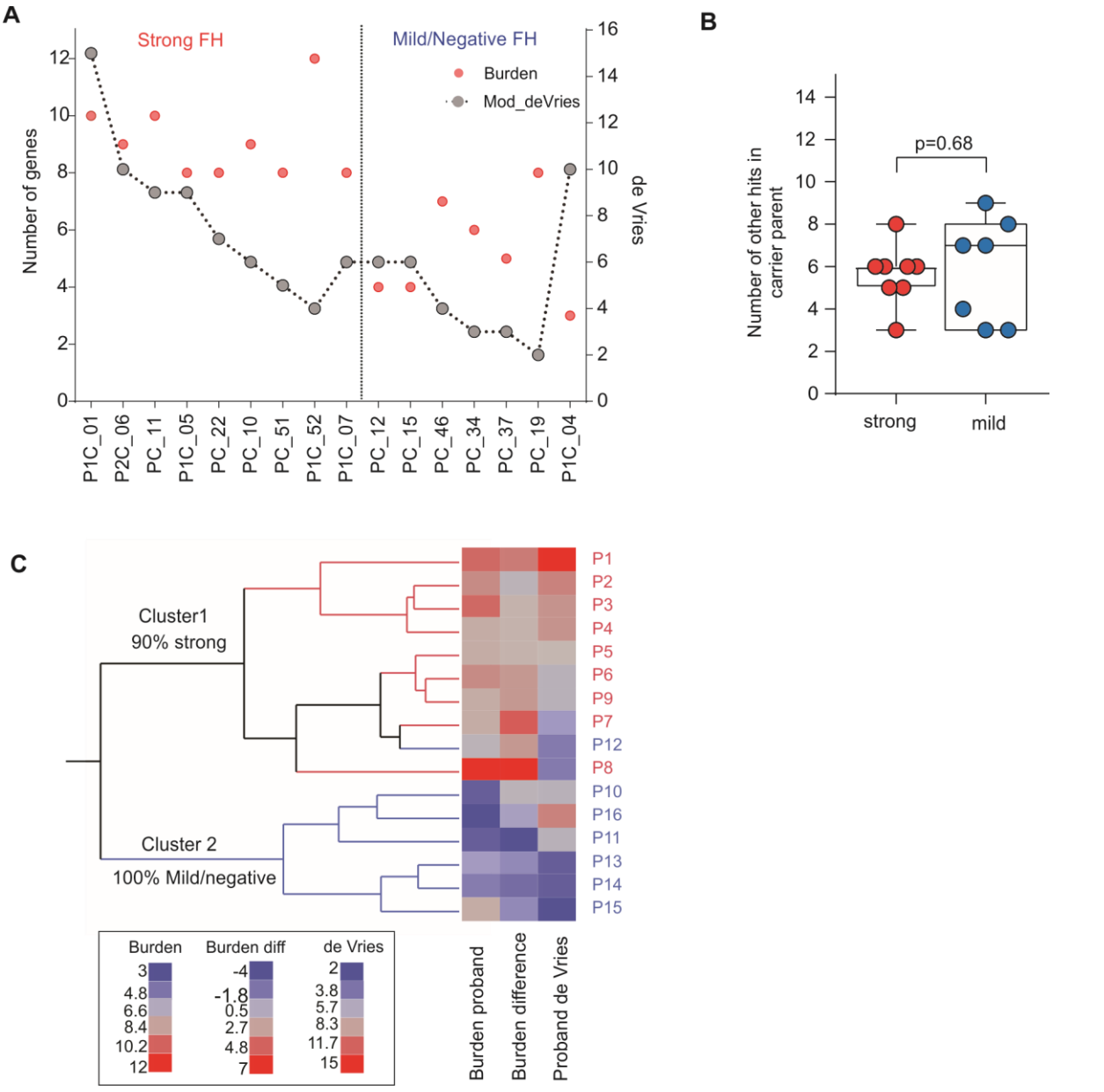


Figure S4. Family history of neurodevelopmental and psychiatric disease is associated with a higher burden of other hits and disease heterogeneity in probands with 16p12.1 deletion. **A.** Burden of other hits (left y-axis, red dots) and clinical severity scores (right y-axis, grey dots) in 16 probands with 16p12.1 deletion from families with strong (n=9) or mild/negative history (n=7) of neurodevelopmental and psychiatric disease is shown. **B.** No difference in the burden of other hits was observed among carrier parents from families with strong family history compared to those with mild/negative family history (one-tailed Mann-Whitney, p=0.68). **C.** Heat map representing hierarchical clustering (Ward method) of probands based on clinical severity scores and burden of other hits identified two clusters of probands that remarkably differed by strong (Cluster I) or mild family history of neuropsychiatric disease (Cluster II).

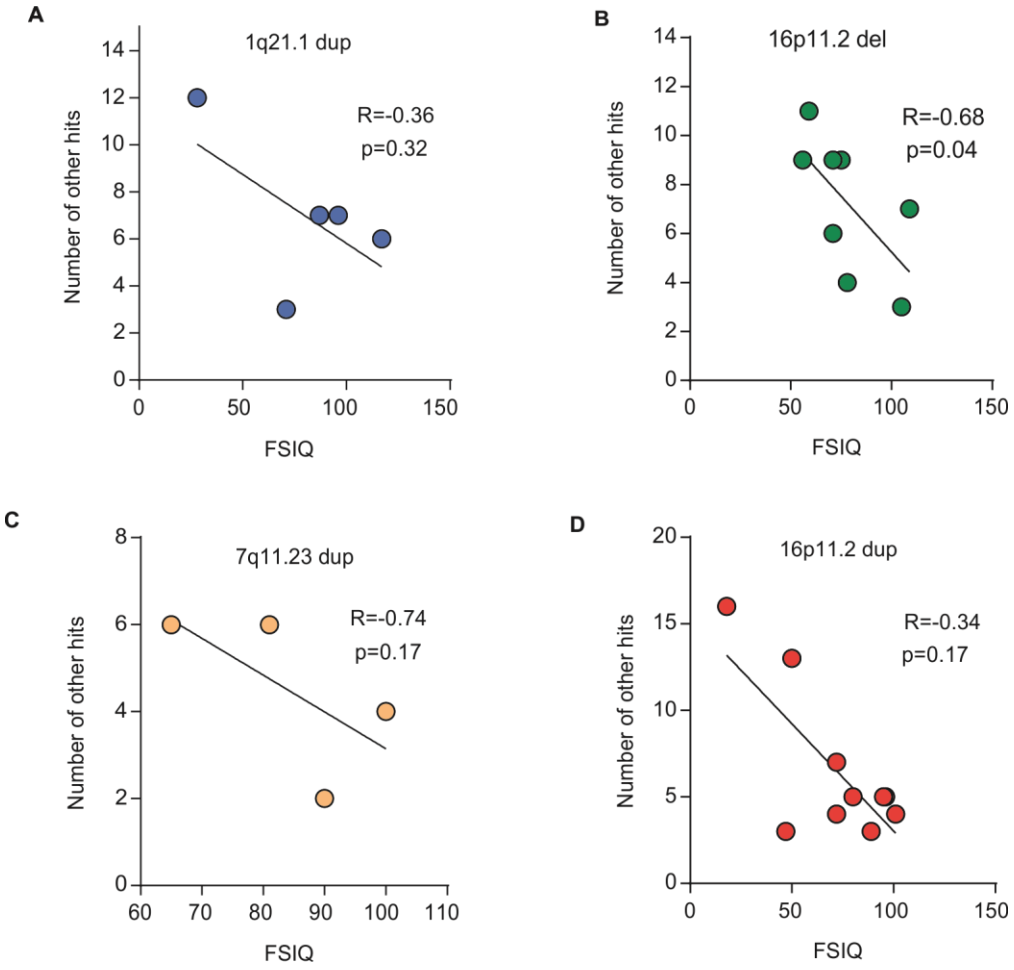


Figure S5. Other-hit burden correlates with FSIQ in SSC cohort individuals with rare CNVs associated with neurodevelopmental disease. The number of other hits correlates (using Spearman correlation measures) with FSIQ scores in individuals with A. 1q21.1 duplication (n=5, R =-0.36, p=0.32), B. 16p11.2 BP4-BP5 deletion (n=8, R=-0.68, p=0.04), C. 16p11.2 BP4-BP5 duplication (n=10, R =-0.74, p=0.17), and D. 7q11.23 duplication (n=4, R =-0.34, p=0.17).

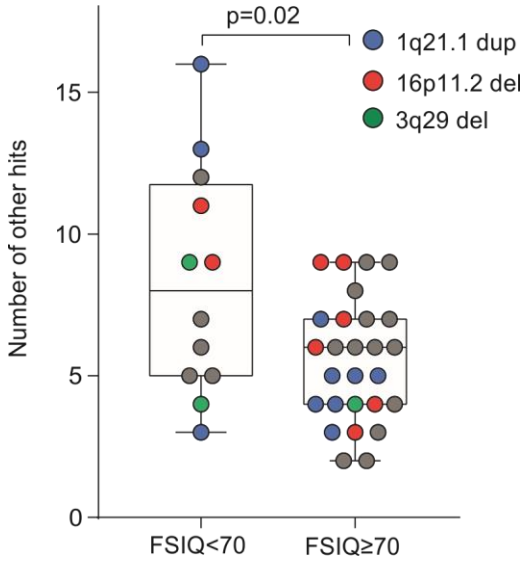


Figure S6. Number of other hits correlates with cognitive phenotypes in carriers of CNVs associated with neurodevelopmental phenotypes. SSC probands with intellectual disability (FSIQ<70, n=12) who carried genomic variants associated with neurodevelopmental disease presented a higher burden of other hits compared to those with FSIQ≥70 (n=27, one-tailed Mann-Whitney p=0.02). Only individuals with CNVs represented in both categories of FSIQ were analyzed. Individuals with 1q21.1 duplication, 16p11.2 deletion and 3q29 deletion are highlighted in the boxplot. Other CNVs are shown in grey.

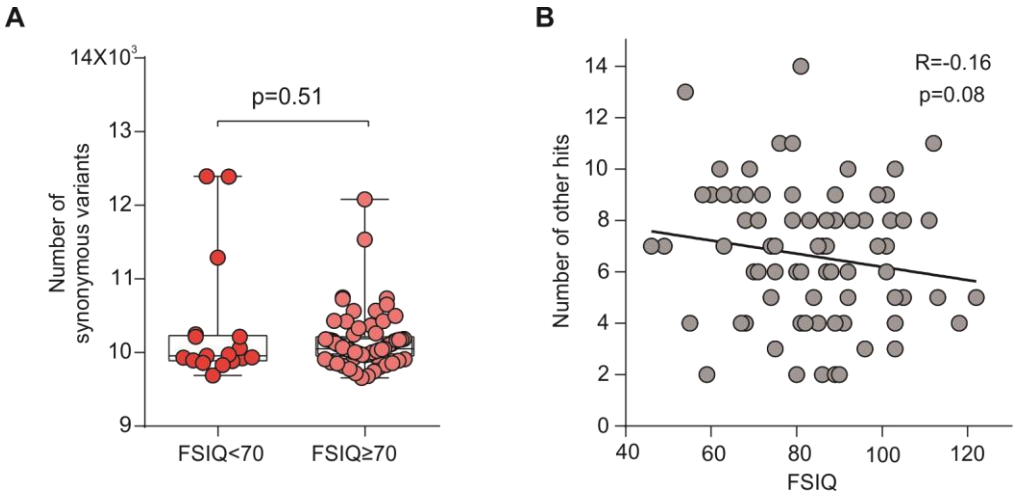


Figure S7. Rare variants in the genetic background and FSIQ scores in SVIP cohort probands with 16p11.2 deletion. A. No difference in the number of synonymous variants in probands with 16p11.2 deletion from the SVIP cohort with (n=17) and without intellectual disability (n=65, two-tailed Mann-Whitney, p=0.51) was observed. **B.** Mild but non-significant correlation was observed between the number of other hits and FSIQ scores in probands with 16p11.2 deletion (n=82, Pearson coefficient, R=-0.16, p=0.08).

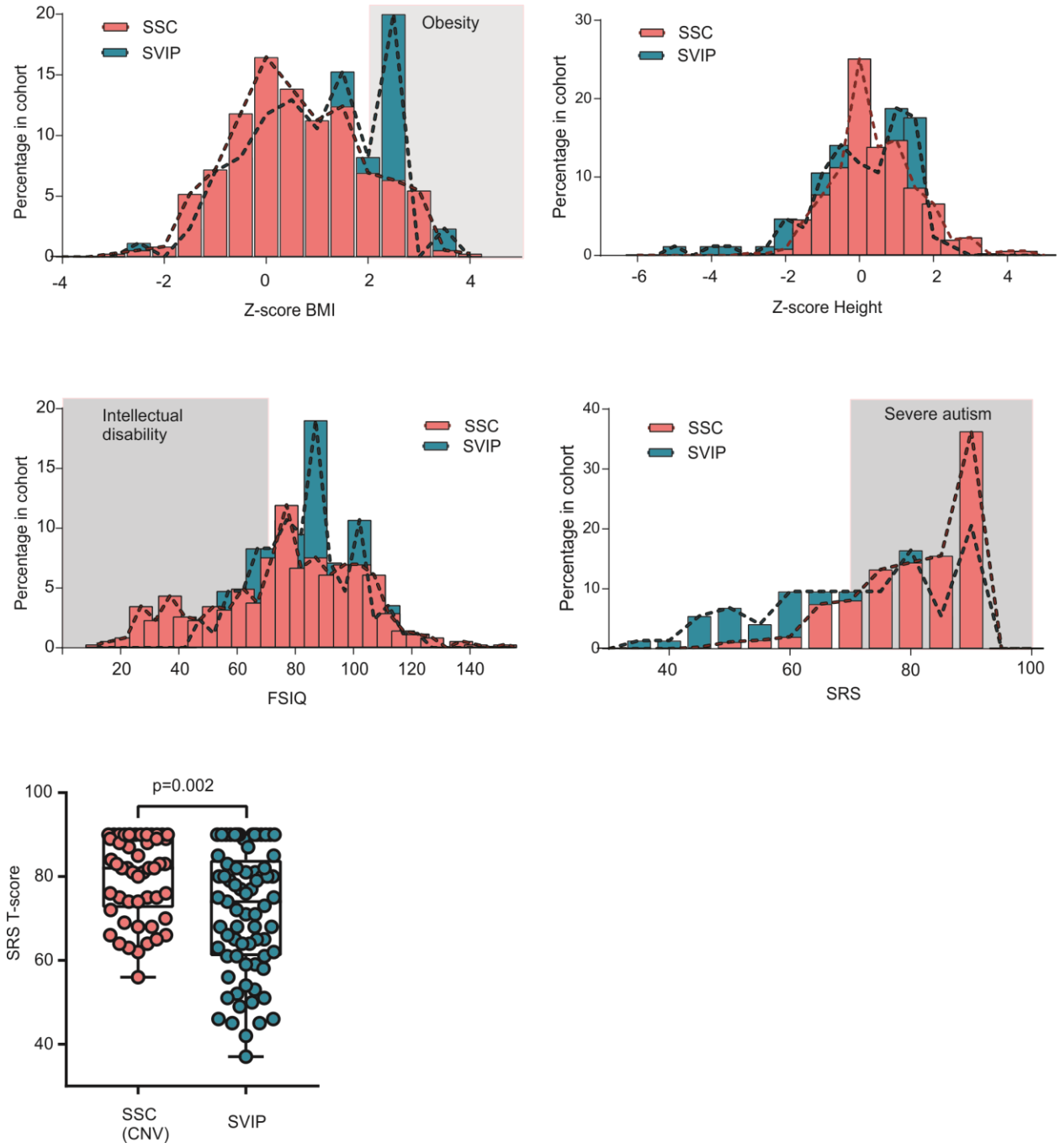


Figure S8. Phenotypic ascertainment differences among probands from the SSC and SVIP cohorts. Distribution of quantitative phenotypes in probands from the SSC cohort carrying pathogenic CNVs (n=53) or *de novo* disruptive variants (n=295) compared to probands with 16p11.2 deletion from the SVIP cohort (n=86). Histograms represent the distribution in each cohort of **A.** BMI z-scores (obesity, BMI>2); **B.** Height z-scores; **C.** FSIQ (intellectual disability, FSIQ<70); and **D.** SRS T-scores (severe autism, SRS T-score>75). **E.** Significant difference in SRS T-scores was observed between probands from the SSC cohort carrying pathogenic CNVs

(including 16p11.2 deletion) and probands from the SVIP cohort with 16p11.2 deletion (two-tailed Mann-Whitney, p=0.002), suggesting ascertainment differences between the two cohorts.

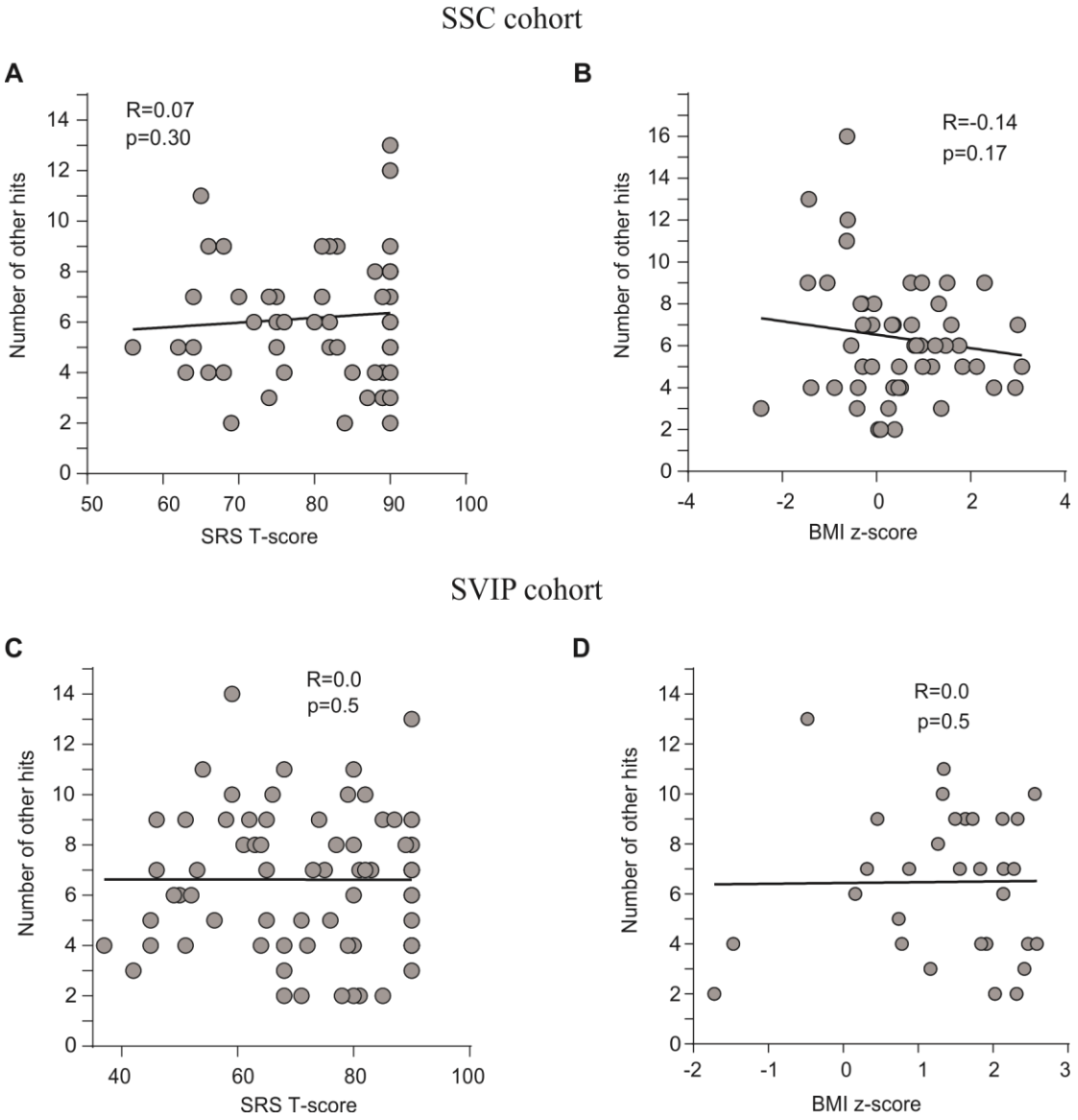


Figure S9. Number of other hits does not correlate with BMI or SRS T-scores in probands with rare CNVs (SSC cohort) or 16p11.2 deletion (SVIP cohort). A. No correlation was found between the number of other hits and SRS T-scores (Pearson correlation, R=0.07, p=0.30) or **B.** BMI z-scores (Pearson correlation, R=-0.14, p=0.17) in probands from the SSC cohort with rare CNVs associated with neurodevelopmental phenotypes (n=53). **C.** SRS T-scores do not correlate (Pearson correlation coefficient R=0.0, p=0.5) with the number of other hits in SVIP probands with the 16p11.2 deletion. **D.** No correlation was found between BMI and other hits among probands with 16p11.2 deletion (SVIP) ≥10 years old (Pearson R=0, p=0.5), the age when the obesity phenotype manifests among carriers of the deletion.

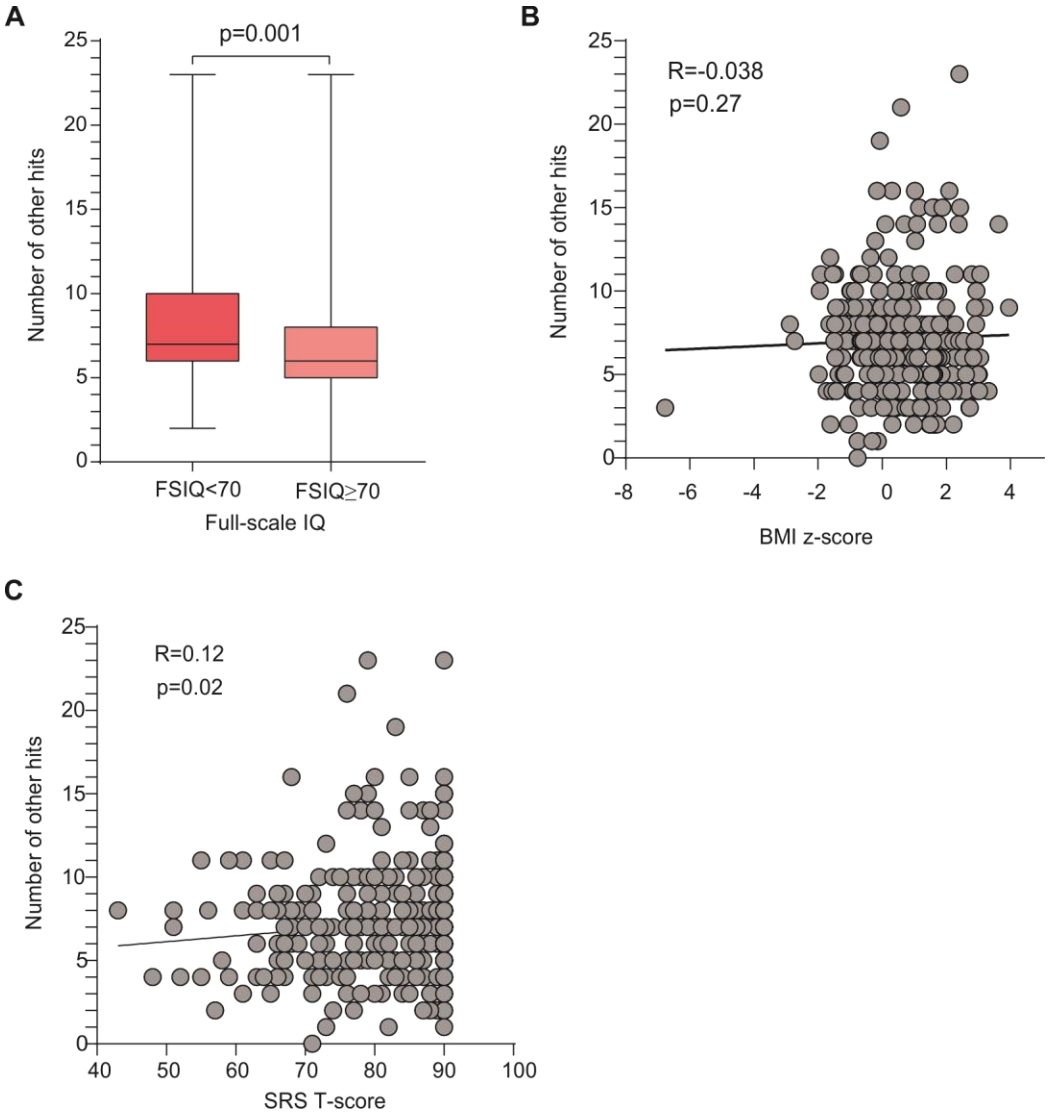


Figure S10. Number of rare variants in the genetic background in probands with *de novo* disruptive variants correlates with FSIQ and SRS T-scores. A. SSC Probands with *de novo* disruptive variants and intellectual disability (FSIQ<70, n=93) presented a higher burden of other hits compared to those with *de novo* disruptive variants and FSIQ≥70 (n=197, one-tailed Mann-Whitney, p=0.001) **B.** No correlation was found in probands with *de novo* disruptive variants between the number of other hits and BMI z-scores (n=275, Spearman correlation, R=-0.038, p=0.27). **C.** A mild correlation was observed between the SRS T-scores (n=295, Spearman correlation, R=0.12, p=0.02) and the number of other hits in SSC probands with *de novo* disruptive variants.

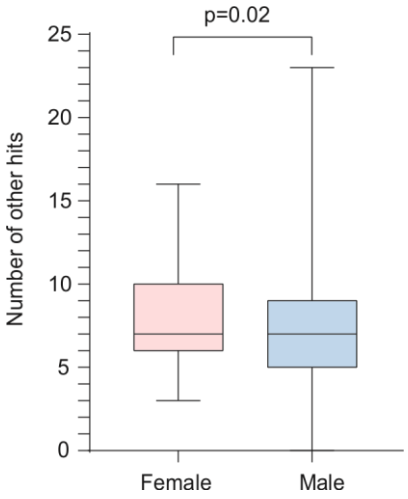


Figure S11. Higher other-hit burden in female probands with *de novo* disruptive variants compared to male probands. An increased number of other hits in female probands with *de novo* disruptive variants (n=46) was observed compared to male probands (n=245, one-tailed Mann-Whitney, p=0.02).

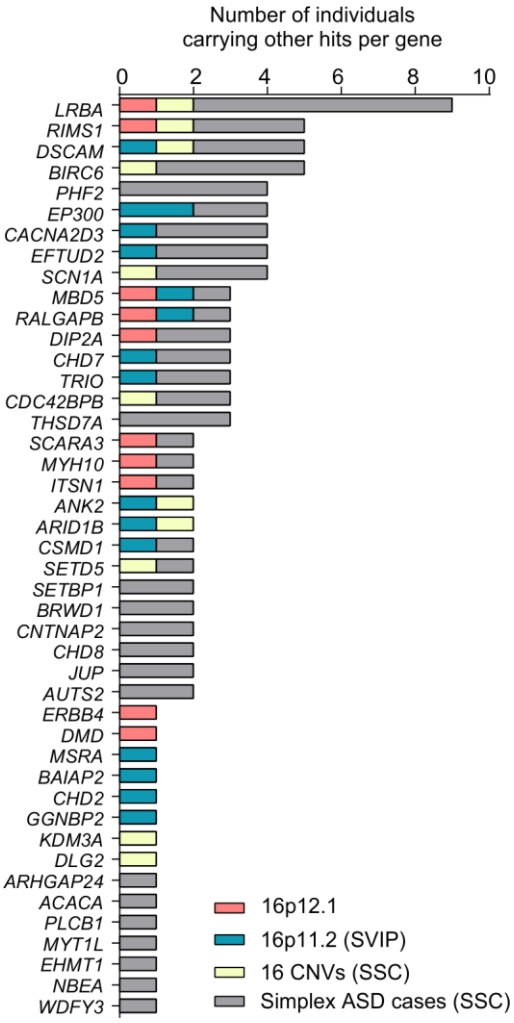


Figure S12. Known neurodevelopmental genes carrying other hits in probands with pathogenic CNVs or *de novo* disruptive variants. This figure shows the number of probands carrying other hits within known neurodevelopmental genes. The data includes individuals with 16p12.1 deletion (recruited in this study, red), 16p11.2 deletion (SVIP cohort, blue), 16 rare CNVs associated with neurodevelopmental phenotypes (SSC cohort, yellow) and de novo disruptive variants in disease-associated genes (SSC cohort, grey).

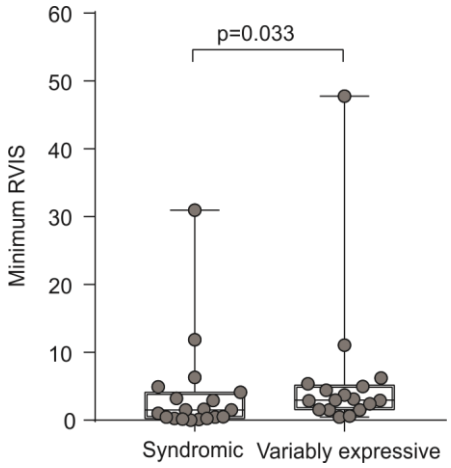


Figure S13. Syndromic CNVs present lower RVIS scores compared to variably expressive CNVs. The minimum RVIS scores for genes within syndromic CNVs are lower than minimum scores for genes within variably expressive CNVs (one-tailed Mann-Whitney, p=0.033).

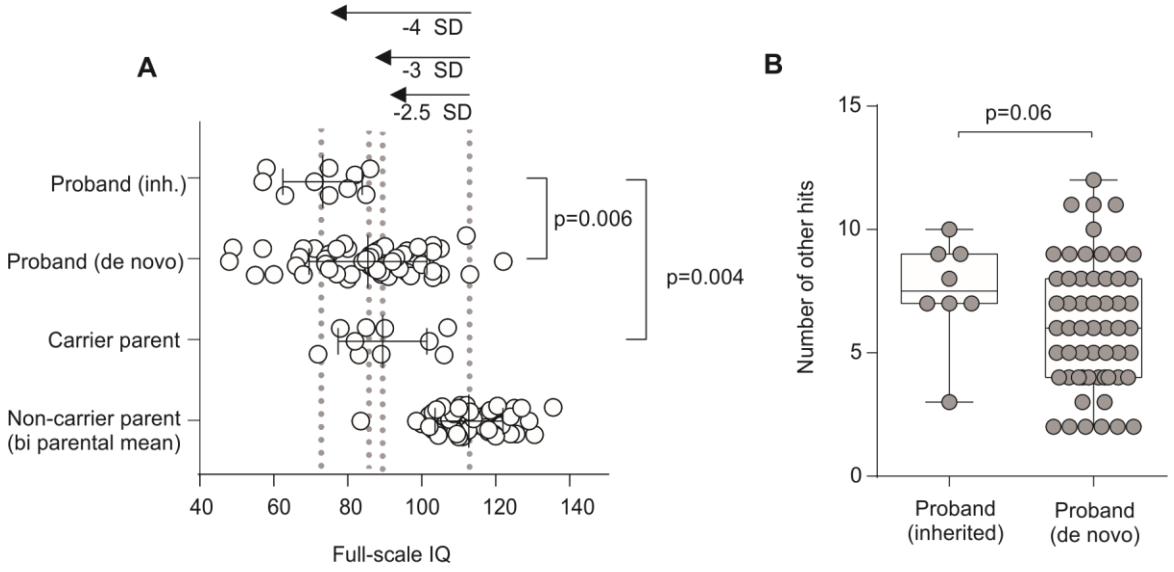


Figure S14. Differences in FSIQ and number of other hits between probands with inherited and *de novo* 16p11.2 deletion. A. SVIP probands with inherited 16p11.2 deletions (n=10) present lower FSIQ scores than those with *de novo* deletions (n=56, one-tailed Mann-Whitney, p=0.006). Probands with *de novo* deletions have a 3SD reduction in FSIQ while carrier parents have a milder 2.5SD reduction in FSIQ compared to the bi-parental mean of unaffected non-carrier parents of probands with *de novo* 16p11.2 deletion. In contrast, probands with an inherited 16p11.2 deletion have a 4SD reduction in FSIQ scores. **B.** Probands with inherited 16p11.2 deletions (n=8) have a non-significant increase in the number of other hits (one-tailed Mann-Whitney, p=0.06) compared to probands with *de novo* 16p11.2 deletions (n=57).

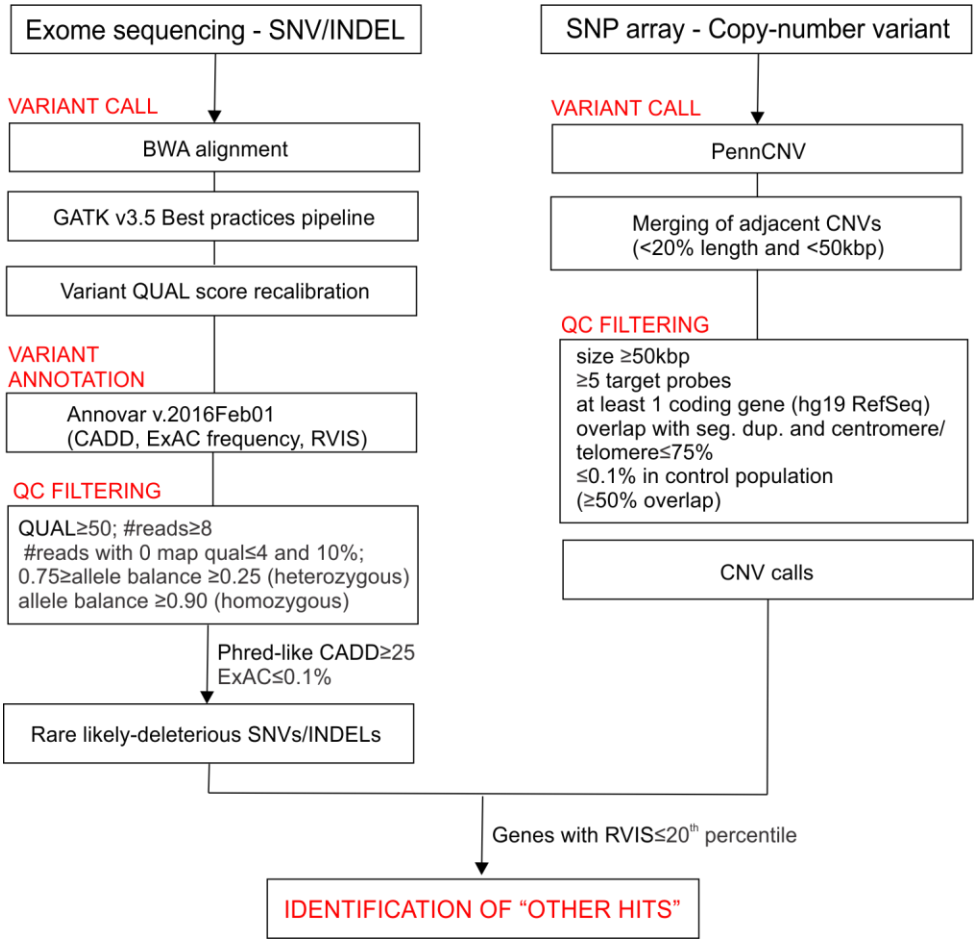


Figure S15. Pipeline for the identification of other hits from exome sequencing and SNP array data.

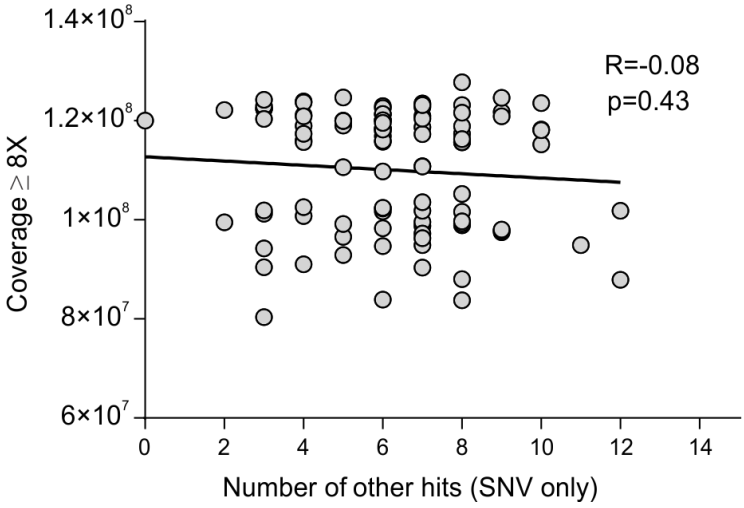


Figure S16. No correlation between the length of genome with coverage ≥8X and the number of other-hit SNVs for individuals with 16p12.1 deletion. No correlation (Pearson correlation, R=-0.08, p=0.43) was found between the length of the genome with coverage ≥8X (minimum coverage used for calling variants) and the number of other hits identified in exome sequencing of 105 individuals from the 16p12.1 deletion cohort.

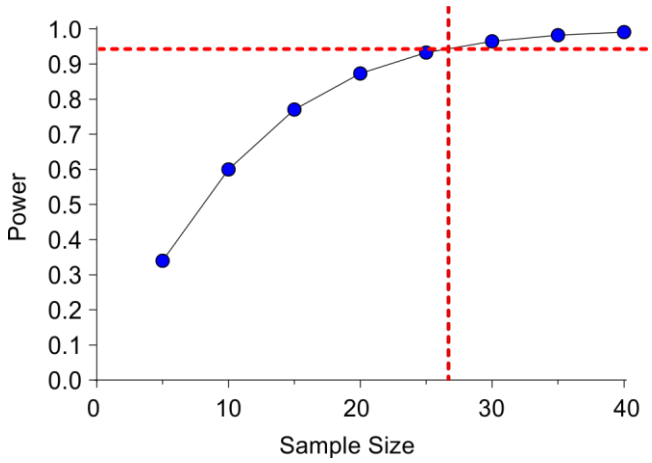


Figure S17. Sample size calculation for 16p12.1 cohort. The number of proband-carrier parent pairs needed to detect a significant change in burden of other hits between probands and carrier parents at power ≥0.8 is n=17 (one-tailed pairwise test). At n=26 pairs in our sample, the one-tailed power is 0.9406.

Supplementary Tables

Table S1. Modified de Vries scoring rubric for uniform quantification of clinical heterogeneity and severity

PHENOTYPES	SCORE
Dysmorphic facial features	max 2
Microcephaly/macrocephaly	1
Growth	max 1
FTT/IUGR/short stature	1
Tall stature	1
Obesity	1
Developmental delay/ motor delay/ speech delay/intellectual disability	max 2
Mild-moderate	1
Severe	2
Abnormal behaviors	max 4
ADHD / Sleep disturbance	1
Schizophrenia	1
Aggression	1
Autism	1
Hypotonia/Hypertonia	1
Epilepsy/Seizures	1
Congenital anomalies	max 3
MRI/brain abnormalities	1
Kidney and urinary tract defects	1
musculoskeletal features	1
cardiac defects	1
genital problems	1
cataracts	1
hearing loss	1
coloboma	1
cleft lip and/or palate	1
Family history	max 3
Father	1
Mother	1
Full sibling(s)	1
	Max 18

Table S2. A summary of inheritance data for 16p12.1 deletion

Ascertainment	Number of individuals
Direct recruitment	
Maternal	34
Paternal	27
De novo	6
Previous studies[1,41]	
Maternal	19
Paternal	5
De novo	1
Combined	
Maternal	53
Paternal	32
De novo	7
Total	92
% Maternal inheritance	57.61*
% Paternal inheritance	34.78 *
% *De novo*	7.60

*Binomial test, p=0.01

Table S3. Phenotypes in probands with 16p12.1 deletion (Excel file)

Table S4. Phenotypes in carrier and non-carrier parents of probands with 16p12.1 deletion

ID	Status	Learning difficulties in school	Depression	ADHD	Alcohol/ Drug abuse	Hallucinations/ Delusions	Seizures	Bipolar disorder
FC_01	FC	Y	Y	Y	N	N	N	N
MC_04	MC	N	N	N	N	N	N	N
M2C_06	MC	Y	Y	Y	Y	Y	N	N
M1C_07	MC	N	N	N	N	N	N	N
M2C_07	MC	Y	Y	Y	N	Y	N	N
MC_10	MC	N	N	N	N	N	N	N
FC_11	FC	N	N	N	N	N	N	N
FC_12	FC	Y	N	N	N	N	Y	N
MC_13	MC	N	N	N	N	N	N	N
MC_15	MC	N	Y	N	Y	N	N	N
FC_16	FC	N	N	N	N	N	N	N
SG260	FC	Y	N	Y	N	N	N	N
MC_22	MC	Y	N	N	N	N	N	N
FC_52	FC	N	Y	N	N	N	N	Y
FC_34	FC	N	N	N	N	N	Y	N
SG100_FC	FC	Y	ND	ND	ND	ND	ND	ND
SG107_FC	FC	Y	ND	ND	ND	ND	ND	ND
SG05_MC	MC	ND	Y	ND	ND	ND	ND	Y
SG06_MC	MC	Y	ND	ND	ND	ND	ND	ND
SG07_MC	MC	Y	ND	ND	ND	ND	Y	ND
SG08_MC	MC	Y	ND	ND	ND	ND	ND	ND
SG09_MC	MC	Y	ND	ND	ND	ND	ND	ND
SG12_FC	FC	Y	ND	ND	ND	ND	ND	ND
SGA3_FC	FC	ND	ND	ND	ND	ND	Y	ND
SGA5_MC	MC	Y	Y	ND	ND	ND	ND	ND
53758-2	FC	N	ND	ND	ND	ND	ND	ND
59152-3	MC	N	ND	ND	ND	ND	ND	ND
44143-2	FC	N	ND	ND	ND	ND	ND	ND
54015-3	MC	ND	ND	ND	ND	ND	ND	ND
32051-3	MC	Y	ND	ND	ND	ND	ND	ND
44098-3	MC	N	ND	ND	ND	ND	ND	ND
59168-3	MC	ND	ND	ND	ND	ND	ND	ND
56411-3	MC	Y	Y	ND	ND	ND	ND	ND
59152-3	MC	N	ND	ND	ND	ND	ND	ND
62244-2	FC	ND	Y	ND	ND	Y	ND	ND
60362-3	MC	N	ND	ND	ND	ND	ND	ND
60450-3	MC	Y	Y	ND	ND	ND	ND	ND
53808-2	FC	ND	ND	ND	ND	ND	ND	ND
57906-3	MC	ND	Y	Y	ND	ND	Y	ND
MNC_01	MNC	Y	Y	N	N	N	N	N
FNC_02	FNC	Y	N	N	N	N	N	N

ID	Status	Learning difficulties in school	Depression	ADHD	Alcohol/ Drug abuse	Hallucinations/ Delusions	Seizures	Bipolar disorder
MNC_02	MNC	N	Y	N	N	N	N	Y
FNC_04	FNC	N	N	N	N	N	N	N
FNC_05	FNC	Y	N	N	N	N	N	N
F1NC_07	FNC	Y	N	N	N	N	N	N
F2NC_07	FNC	Y	Y	Y	Y	Y	N	N
FNC_10	FNC	N	N	N	N	N	N	N
MNC_11	MNC	N	N	N	N	N	N	N
MNC_12	MNC	N	Y	N	N	N	N	N
FNC_13	FNC	N	N	N	N	N	N	N

"FC"= Father carrier, "MC"= Mother carrier, "FNC"= Father non-carrier, "MNC"= Mother non-carrier, "Y"= Yes, "N" = No, "ND" = Not determined.

Table S5. Summary of exome sequencing and SNP arrays performed on 16p12.1 probands and family members

		Number of individuals
Complete families		24
	3-generation families	3
	Trios	17
	Quads	4
Incomplete families		2
	Trios	1
	Quads	1
All individuals (total)		105
	16p12.1 del carriers	59
	Non-16p12.1 del carriers	46
	Carrier children	33
	Carrier children male	21
	Carrier children female	12
	Carrier fathers	10
	Carrier mothers	13
	Carrier grandparents	3
	Non-carrier fathers	16
	Non-carrier mothers	14
	Other non-carrier family members	16
	De novo cases	3

Table S6. Coding variants in 16p12.1 genes identified on the non-deleted chromosome of probands with 16p12.1 deletion

CHR	POS	REF	ALT	Patient ID	Observed in a carrier parent in the cohort	Gene	CADD score	ExAC freq.	dbSNP ID
chr16	21976762	G	A	PC_11	Yes	*UQCRC2*	15.56	0.0429	rs4850
chr16	22319517	T	G	PC_49	Yes	*POLR3E*	13.48	0.0613	rs2347
chr16	22319517	T	G	PC_51	Yes	*POLR3E*	13.48	0.0613	rs2347
chr16	22149688	T	C	PC_02	Yes	*VWA3A*	15.57	0.0132	rs145806753
chr16	22149688	T	C	PC_34	Yes	*VWA3A*	15.57	0.0132	rs145806753
chr16	22157582	C	A	PC_11	Yes	*VWA3A*	10.05	0.0424	rs61744122
chr16	22237273	C	G	P1C_01	No	*EEF2K*	17.1	0.1083	rs17841292
chr16	22019646	AG	A	P2C_07	Yes	*C16orf52*	NA	0.144	rs201044196
chr16	22019646	AG	A	PC_19	Yes	*C16orf52*	NA	0.144	rs201044196
chr16	22019646	AG	A	PC_22	Yes	*C16orf52*	NA	0.144	rs201044196
chr16	22019646	AG	A	PC_33	Yes	*C16orf52*	NA	0.144	rs201044196
chr16	22019646	AG	A	PC_49	Yes	*C16orf52*	NA	0.144	rs201044196
chr16	22092067	A	C	PC_11	Yes	*C16orf52*	11.54	0.0253	rs72784938

Table S7. *De novo* variants identified in probands with 16p12.1 deletion and carrier siblings

ID	CHR	POS	REF	ALT	Gene	Variant Type	CADD score	ExAC freq.	dbSNP ID
P1C_01	chr17	12887918	C	T	*ARHGAP44*	synonymous	3.784	4.5E-05	NA
P1C_01	chr3	9488832	T	TAC	*SETD5*	frameshift	NA	NA	NA
PC_02	chr9	20789546	G	T	*FOCAD*	nonsynonymous	20.7	NA	NA
P1C_04	chr11	3716731	G	A	*NUP98*	nonsynonymous	13.19	NA	NA
P2C_04	chr1	109271413	G	T	*FNDC7*	nonsynonymous	0.143	NA	NA
P2C_04	chr8	87460677	A	C	*WWP1*	synonymous	9.482	NA	NA
P1C_05	chr19	758712	C	T	*MISP*	nonsynonymous	13.9	NA	NA
P2C_07	chr17	80012469	G	A	*GPS1*	nonsynonymous	15.92	NA	NA
PC_12	chr3	113344957	G	A	*SIDT1*	synonymous	5.274	1.6E-05	NA
PC_12	chr4	95575739	A	G	*PDLIM5*	nonsynonymous	21.9	NA	NA
PC_19	chr18	19075644	A	G	*GREB1L*	nonsynonymous	5.19	NA	NA
PC_19	chr2	231949783	T	A	*PSMD1*	stopgain	42	NA	NA
PC_20	chr8	37690603	C	T	*ADGRA2*	synonymous	0.896	NA	Yes*
PC_21	chr6	126075667	G	C	*HEY2*	nonsynonymous	21.7	NA	NA
PC_37	chr1	153721226	G	C	*INTS3*	nonsynonymous	12.09	NA	NA

* rs372033560

Table S8. Private and rare gene-disruptive variants in disease-associated genes identified in probands with 16p12.1 deletion.

ID	CHR	POS	REF	ALT	Gene	Variant Type	CADD score	ExAC freq.	dbSNP ID
P1C_01	chr3	9488832	T	TAC	*SETD5*	Frameshift insertion	NA	NA	NA
P1C_01	chr9	133927967	C	T	*LAMC3*	stopgain	36	NA	NA
P1C_01	chrX	33229421	C	T	*DMD*	stopgain	38	NA	NA
PC_02	chr2	212483904	C	G	*ERBB4*	nonsynonymous	26	NA	NA
PC_02	chr16	58608972	G	T	*CNOT1*	nonsynonymous	33	NA	NA
P1C_05	chr11	47290147	G	A	*NR1H3*	nonsynonymous	36	0.0002	rs61731956
P1C_05	chr16	87417050	G	A	*FBXO31*	nonsynonymous	27.3	2E-05	NA
P1C_06	chr11	103124071	G	A	*DYNC2H1*	nonsynonymous	34	4E-05	NA
P1C_06	chr15	63978674	G	C	*HERC1*	nonsynonymous	28.2	2E-05	rs375968062
P1C_06	chr4	38020008	G	A	*TBC1D1*	nonsynonymous	35	2E-05	NA
P2C_06	chr22	19423166	G	A	*MRPL40*	nonsynonymous	26.7	3E-05	NA
P2C_06	chr4	38020008	G	A	*TBC1D1*	nonsynonymous	35	2E-05	NA
P1C_07	chr11	58891938	A	ACT	*FAM111B*	Frameshift insertion	NA	9E-05	NA
P1C_07	chr17	10354712	T	C	*MYH4*	nonsynonymous	34	NA	NA
P2C_07	chr2	220284873	G	A	*DES*	nonsynonymous	35	0.0002	rs144261171
P2C_07	chr11	58891938	A	ACT	*FAM111B*	Frameshift insertion	NA	9E-05	NA
P2C_07	chr11	58891938	A	ACT	*FAM111B*	Frameshift insertion	NA	9E-05	NA
PC_10	chr12	91363873	A	T	*EPYC*	stopgain	36	0.0001	rs150809530
PC_10	chr15	24923851	CAG	C	*NPAP1*	Frameshift deletion	NA	NA	NA
PC_10	chr1	19465697	C	T	*UBR4*	nonsynonymous	35	NA	NA
PC_10	chr19	44153044	A	G	*PLAUR*	stoploss	13.02	0.0002	rs140046361
PC_11	chr2	157425430	A	G	*GPD2*	nonsynonymous	26.9	8E-06	NA
PC_11	chr6	75901477	C	G	*COL12A1*	NA	15.44	8E-06	NA
PC_12	chr7	103197510	G	A	*RELN*	nonsynonymous	27.6	0.0002	rs114190729
PC_13	chr6	157099425	A	AG C	*ARID1B*	Frameshift insertion	NA	0.0004	NA
PC_13	chr4	151357910	G	T	*LRBA*	nonsynonymous	32	0.0007	rs151286835
PC_18	chr3	173525476	G	A	*NLGN1*	nonsynonymous	25.6	8E-06	rs147780897
PC_18	chr8	27516403	G	A	*SCARA3*	nonsynonymous	26.3	9E-05	rs150905493
PC_19	chr6	56480833	G	A	*DST*	stopgain	48	8E-06	NA
PC_19	chr2	179664292	C	T	*TTN*	nonsynonymous	32	2E-05	rs151174349
PC_19	chr5	112926913	T	A	*YTHDC2*	nonsynonymous	33	0.0002	rs200971375
PC_19	chr11	126316710	C	T	*KIRREL3*	nonsynonymous	26.3	6E-05	rs374559484
PC_19	chr1	210637960	G	A	*HHAT*	nonsynonymous	33	8E-06	NA
PC_20	chr5	130771676	G	A	*RAPGEF6*	stopgain	36	0.0008	rs183985113
PC_20	chr11	119167631	TCTTC	T	*CBL*	Frameshift deletion	NA	3E-05	NA

ID	CHR	POS	REF	ALT	Gene	Variant Type	CADD score	ExAC freq.	dbSNP ID
PC_20	chr19	12805516	CGACA	C	*FBXW9*	Frameshift deletion	NA	2E-05	NA
PC_20	chr2	179406192	G	A	*TTN*	nonsynonymous	28.4	0.0001	NA
PC_20	chr12	50195698	C	T	*NCKAP5L*	nonsynonymous	32	5E-05	rs374734587
PC_21	chr11	126314949	C	T	*KIRREL3*	nonsynonymous	30	3E-05	rs201882059
PC_21	chr17	30190491	C	G	*UTP6*	nonsynonymous	28	2E-05	NA
PC_22	chr13	101763037	G	C	*NALCN*	stopgain	41	NA	NA
PC_22	chr17	79634829	G	T	*CCDC137*	stopgain	36	NA	NA
PC_22	chr19	11557904	GAC	G	*PRKCSH*	Frameshift deletion	NA	NA	NA
PC_22	chr1	202727613	A	G	*KDM5B*	nonsynonymous	32	NA	NA
PC_22	chr19	6361896	C	T	*CLPP*	nonsynonymous	35	9E-06	NA
PC_31	chr5	76709089	C	CG	*PDE8B*	Frameshift insertion	NA	NA	NA
PC_31	chr10	16946057	G	GTT ATA TAA	*CUBN*	stopgain	NA	0.0001	NA
PC_31	chr11	46564518	TG	T	*AMBRA1*	Frameshift deletion	NA	NA	NA
PC_31	chr19	36297980	AG	A	*PRODH2*	Frameshift deletion	NA	0.0004	NA
PC_31	chr20	37177397	C	T	*RALGAPB*	stopgain	48	NA	NA
PC_31	chr2	48063103	G	T	*FBXO11*	nonsynonymous	25.4	NA	NA
PC_31	chr10	60560686	G	A	*BICC1*	nonsynonymous	36	2E-05	NA
PC_33	chr1	39917864	G	A	*MACF1*	nonsynonymous	35	0.0003	rs146089082
PC_33	chr1	230895256	A	G	*CAPN9*	NA	22.5	NA	NA
PC_34	chr17	29549007	GAAA	G	*NF1*	Frameshift deletion	NA	NA	NA
PC_34	chr5	90074814	G	A	*ADGRV1*	nonsynonymous	37	3E-05	rs182452385
PC_37	chr14	64634063	G	A	*SYNE2*	nonsynonymous	34	0.0008	rs149227847
PC_37	chr15	23060834	C	T	*NIPA1*	nonsynonymous	32	NA	NA
PC_37	chr16	58615358	C	T	*CNOT1*	nonsynonymous	26.8	2E-05	rs367777689
PC_37	chr17	76525627	G	A	*DNAH17*	nonsynonymous	27	0.0007	rs201764607
PC_46	chr6	99891524	C	A	*USP45*	stopgain	47	0.0001	rs141844660
PC_46	chr6	72955537	G	A	*RIMS1*	nonsynonymous	33	NA	NA
PC_46	chr11	9048988	C	T	*SCUBE2*	nonsynonymous	35	8E-05	rs370883793
PC_46	chr17	17119793	G	A	*FLCN*	nonsynonymous	27.4	0.0001	rs143183215
PC_48	chr17	58740891	G	A	*PPM1D*	nonsynonymous	28.6	NA	NA
PC_48	chr21	35186357	C	T	*ITSN1*	nonsynonymous	27.9	0.001	rs143723211
PC_48	chr12	124289588	G	C	*DNAH10*	NA	12.6	8E-06	NA
PC_49	chr1	33282806	T	A	*YARS*	nonsynonymous	34	3E-05	NA
PC_49	chr2	149221327	G	A	*MBD5*	nonsynonymous	25	0.0008	rs34995577
PC_49	chr13	24868977	C	T	*SPATA13*	nonsynonymous	36	0.0008	rs140467795
PC_51	chr17	8508223	C	T	*MYH10*	nonsynonymous	26.5	NA	NA
P1C_52	chr20	62729293	GAC	G	*OPRL1*	Frameshift deletion	NA	6E-05	NA

Table S9. Copy-number variants in families with 16p12.1 deletion (Excel file)

Table S10. Canonical pathway analysis of genes with identified other hits in probands and carrier parents with 16p12.1 deletion.

	Probands	
	FDR*	Z-score
Calcium signaling[42]	9.29E-05	-2.2
Corticotropin releasing hormone signaling	1.32E-03	-2.6
Dopamine_DARPP32 feedback in cAMP signaling	3.75E-02	-2.5
GNRH signaling	4.62E-02	-2.5
Wnt/Ca^{++} signaling[43,44]	4.96E-02	-2
CREBB signaling in neurons	4.96E-02	-2.6
	Carrier parents	
	FDR	Z-score
CREBB signaling in neurons	4.89E-03	-2.6
Synaptic long term depression	7.54E-03	-2.5

*One-tailed Fisher's Exact test with Benjamini-Hochberg multiple testing correction was used for statistical analysis.

Table S11. Number of probands from the SSC cohort carrying rare CNVs associated with neurodevelopmental phenotypes

CNV	Number of probands
1q21.1 dup	5
15q11.1-13.1 dup	3
15q13.2,15q13.3 del	3
16p11.2 del	8
16p11.2 dup	10
16p11.2 distal del	1
16p13.11 del	3
16p13.11 dup	3
17q12 del	3
17q11.2 NF1 dup	1
22q11.21 del	1
22q11.21 dup	2
2q23.1 del	2
3q29 del	3
4p16.2-16.3 Wolf-Hirschhorn dup	1
7q11.23 dup	4

Table S12. Inherited disruptive variants in neurodevelopmental disease-associated genes found in autism proband-unaffected sibling pairs in the SSC cohort (Excel file)

Table S13. Genes originally identified to carry *de novo* disruptive variants in autism simplex cases but identified as other hits in this study (in probands with first-hit pathogenic CNVs or SNVs)

Gene	Recurrence as other hit	16p12.1 del (variant type; CADD score)	16p11.2 del (SVIP) (variant type; CADD score)	15 CNVs (SSC) (variant type; CADD score)	Simplex ASD (SSC) (variant type; CADD score)
RIMS1	5	nonsyn;33	0	stopgain;41	nonsyn;26.2/stopgain;41/stopgain;39
MYH2	5	nonsyn;29.7	0	0	nonsyn;33/nonsyn;35/nonsyn;29.1/nonsyn;33
DSCAM	5	0	nonsyn;26.7	nonsyn;27.6	nonsyn;32/nonsyn;27.6/nonsyn;33
ADAMTS9	5	0	0	nonsyn;27.4	nonsyn;33/nonsyn;27.4/nonsyn;27.8/ nonsyn;27
PHF2	4	0	0	0	nonsyn;27/nonsyn;28.9/nonsyn;28.9/nonsyn;28.9
KDM5B	4	nonsyn;32	0	0	nonsyn;26.7/nonsyn;27.3;stopgain;41
CSMD2	4	0	nonsyn;34	0	nonsyn;28.2/nonsyn;34/nonsyn;27.5
SPAG9	4	0	nonsyn;25.5/ nonsyn;25.5	nonsyn;25.5	nonsyn;32
MBD5	3	nonsyn;25	nonsyn;25	0	nonsyn;31
SCUBE2	3	nonsyn;35	0	0	nonsyn;35/nonsyn;28
DIP2A	3	nonsyn;26	0	0	nonsyn;26.7/stopgain;45
RERE	2	0	nonsyn;27.8	0	nonsyn;33
CDAN1	3	0	nonsyn;25.6	0	nonsyn;35/nonsyn;26.8
ACOX2	3	0	stopgain;40	nonsyn;33	nonsyn;26.3
CDC42BPB	3	0	0	nonsyn;35	nonsyn;28/nonsyn;33
THSD7A	3	0	0	0	nonsyn;25.3/nonsyn;25.1/nonsyn;25.6
UBN2	2	nonsyn;34	0	0	nonsyn;25.3
MYH10	2	nonsyn;26.5	0	0	nonsyn;35
ANK2	2	0	nonsyn;25.6	nonsyn;28.7	0
ARID1B	2	0	nonsyn;25	nonsyn;25.4	0
SIK3	2	0	nonsyn;33	nonsyn;25.9	0
HECTD1	2	0	nonsyn;26.9	0	nonsyn;35
DOCK5	2	0	nonsyn;34	0	nonosyn;32
TANC2	2	0	0	nonsyn;27	nonsyn;32
SETBP1	2	0	0	0	nonsyn;34/stopgain;45
SUFU	2	0	0	0	nonsyn;28.8/nonsyn;35
TRIP12	2	0	0	0	nonsyn;29.6/nonsyn27.9
PFKFB2	2	0	0	0	nonsyn29.6/nonsyn29.6
BRWD1	2	0	0	0	nonsyn26.9/nonsyn28.8
GIGYF1	2	0	0	0	nonsyn;32/stopgain;49
EPHB2	2	0	0	0	nonsyn;25.2/nonsyn;35
DSG3	2	0	0	0	stopgain;31/nonsyn;25.9
CHD8	2	0	0	0	stopgain;29.3/stopgain;27.8
TECTA	2	0	0	0	nosnsyn;27/nosnsyn;26.5
XPO4	1	0	0	0	nonsyn;25.6
FBXO11	1	nonsyn;25.4	0	0	0

Gene	Recurrence as other hit	16p12.1 del (variant type; CADD score)	16p11.2 del (SVIP) (variant type; CADD score)	16 CNVs (SSC) (variant type; CADD score)	Simplex ASD (SSC) (variant type; CADD score)
RCBTB1	1	0	stopgain;39	0	0
WDR33	1	0	stopgain;39	0	0
DVL3	1	0	nonsyn;31	0	0
CHD2	1	0	nonsyn;33	0	0
NF1	1	0	nonsyn;35	0	0
ZC3H18	1	0	nonsyn;28.7	0	0
ABTB1	1	0	nonsyn;37	0	0
DOT1L	1	0	nonsyn;31	0	0
PFKL	1	0	0	nonsyn;34	0
LRP6	1	0	0	nonsyn;33	0
DNMT3A	1	0	0	nonsyn;28	0
CHD1	1	0	0	nonsyn;26.3	0
PLCD4	1	0	0	0	stopgain;40
FAM91A1	1	0	0	0	nonsyn;26.3
DHX29	1	0	0	0	nonsyn;29
MYO1E	1	0	0	0	nonsyn;25.6
IGSF3	1	0	0	0	nonsyn;35
MED13	1	0	0	0	nonsyn;25.3
NBEA	1	0	0	0	nonsyn;31
GOPC	1	0	0	0	nonsyn;34
WDFY3	1	0	0	0	nonsyn;26.2
LARP4B	1	0	0	0	nonsyn;28.5

"Recurrence as a other hit" indicates how many times the gene was observed as a other hit in our study.
"Nonsyn" = non-synonymous variant

Table S14. Genes with other hits associated with skeletal, muscular, cardiovascular or renal disease (Excel file)

Table S15. Biological processes enriched among genes carrying other hits in SSC probands with *de novo* gene-disruptive variants

PANTHER GO-Slim Biological Process	Expected number	Observed number	FDR
sensory perception of sound (GO:0007605)	31	15	5.09E-07
negative regulation of apoptotic process (GO:0043066)	98	26	3.56E-07
muscle contraction (GO:0006936)	112	20	8.84E-03
anatomical structure morphogenesis (GO:0009653)	470	80	6.18E-13
cellular component morphogenesis (GO:0032989)	339	57	1.03E-08
transmembrane receptor protein tyrosine kinase signaling pathway (GO:0007169)	151	24	9.82E-03
cellular component movement (GO:0006928)	407	61	1.77E-07
cell adhesion (GO:0007155)	336	47	1.07E-04
biological adhesion (GO:0022610)	336	47	1.07E-04
regulation of catalytic activity (GO:0050790)	338	47	1.25E-04
regulation of molecular function (GO:0065009)	412	56	2.19E-05
regulation of phosphate metabolic process (GO:0019220)	479	63	1.12E-05
cell differentiation (GO:0030154)	459	60	2.85E-05
homeostatic process (GO:0042592)	262	34	1.68E-02
intracellular signal transduction (GO:0035556)	1000	123	3.09E-10
MAPK cascade (GO:0000165)	329	40	1.64E-02
apoptotic process (GO:0006915)	440	49	2.26E-02
developmental process (GO:0032502)	1569	173	6.54E-11
vesicle-mediated transport (GO:0016192)	800	84	8.08E-04
single-multicellular organism process (GO:0044707)	1463	149	7.71E-07
multicellular organismal process (GO:0032501)	1479	149	1.54E-06
intracellular protein transport (GO:0006886)	688	68	3.89E-02
protein transport (GO:0015031)	720	70	4.99E-02
catabolic process (GO:0009056)	1127	108	1.89E-03
phosphate-containing compound metabolic process (GO:0006796)	858	82	2.56E-02
transport (GO:0006810)	1668	159	1.60E-05
cell communication (GO:0007154)	2509	238	6.31E-09
localization (GO:0051179)	1970	186	2.04E-06
signal transduction (GO:0007165)	2162	198	6.04E-06
cellular process (GO:0009987)	7905	647	2.09E-16
nucleobase-containing compound metabolic process (GO:0006139)	2050	167	4.90E-02
primary metabolic process (GO:0044238)	4346	345	8.12E-05
metabolic process (GO:0008152)	5416	427	2.85E-06

Table S16. Biological processes enriched among genes with other hits in SVIP probands with 16p11.2 deletion

PANTHER GO-Slim Biological Process	Expected number	Observed number	FDR
sensory perception of sound (GO:0007605)	31	12	3.42E-09
muscle contraction (GO:0006936)	112	14	1.08E-04
nervous system development (GO:0007399)	238	20	2.18E-04
cellular component morphogenesis (GO:0032989)	339	27	8.13E-06
anatomical structure morphogenesis (GO:0009653)	470	37	3.10E-08
cell adhesion (GO:0007155)	336	26	2.54E-05
biological adhesion (GO:0022610)	336	26	2.54E-05
mitosis (GO:0007067)	221	17	4.33E-03
mesoderm development (GO:0007498)	278	18	2.22E-02
system development (GO:0048731)	357	22	8.01E-03
cell differentiation (GO:0030154)	459	28	8.48E-04
cellular component movement (GO:0006928)	407	23	1.95E-02
intracellular signal transduction (GO:0035556)	1000	53	3.79E-06
developmental process (GO:0032502)	1569	80	3.11E-09
single-multicellular organism process (GO:0044707)	1463	68	5.26E-06
multicellular organismal process (GO:0032501)	1479	68	7.95E-06
neurological system process (GO:0050877)	816	36	3.63E-02
system process (GO:0003008)	910	39	3.49E-02
cell communication (GO:0007154)	2509	100	4.37E-06
signal transduction (GO:0007165)	2162	86	6.01E-05
localization (GO:0051179)	1970	70	3.34E-02
cellular process (GO:0009987)	7905	255	5.72E-10

Table S17. Syndromic and variably expressive CNVs associated with neurodevelopmental disorders

Region	Effect of deletion	Chr	hg19 start	hg19 end	hg19 length
1p36	syndromic	1	10,001	10,077,413	10,067,412
2q37	syndromic	2	239,705,243	242,471,327	2,766,084
Wolf-Hirschhorn	syndromic	4	1,529,198	2,030,202	501,004
Sotos	syndromic	5	175,717,394	177,057,394	1,340,000
6q16	syndromic	6	100,813,279	100,943,279	130,000
Williams	syndromic	7	72,742,064	74,142,064	1,400,000
8p23.1	syndromic	8	8,092,590	11,892,591	3,800,001
9q34	syndromic	9	137,810,179	141,080,179	3,270,000
Prader-Willi/Angelmans	syndromic	15	24,818,907	28,426,405	3,607,498
15q24	syndromic	15	72,952,946	74,362,947	1,410,001
15q24.2q24.3	syndromic	15	75,972,945	78,202,945	2,230,000
Rubinstein-Taybi	syndromic	16	3,749,999	3,949,999	200,000
17p13.3	syndromic	17	1,053,250	2,633,250	1,580,000
Smith-Magenis	syndromic	17	16,789,275	18,299,275	1,510,000
NF1	syndromic	17	29,095,874	30,275,887	1,180,013
17q21.31	syndromic	17	43,704,217	44,164,182	459,965
19p13.12	syndromic	19	13,079,000	16,699,000	3,620,000
Phelan-McDermid	syndromic	22	43,000,056	51,163,134	8,163,078
PLP1	syndromic	X	102,413,344	113,413,741	11,000,397
1q21.1	Variably expressive	1	146,573,376	147,393,376	820,000
2q23.1	Variably expressive	2	148,723,530	149,293,530	570,000
3q29	Variably expressive	3	195,745,603	197,355,603	1,610,000
6p25	Variably expressive	6	155,000	6,055,001	5,900,001
10q23	Variably expressive	10	81,960,020	88,800,020	6,840,000
15q11.2	Variably expressive	15	22,798,636	23,088,559	289,923
15q13.3	Variably expressive	15	31,132,708	32,482,708	1,350,000
15q25	Variably expressive	15	83,182,945	84,738,996	1,556,051
16p13.11	Variably expressive	16	15,502,499	16,292,499	790,000
16p12.1	Variably expressive	16	21,942,499	22,432,499	490,000
16p11.2 distal	Variably expressive	16	28,822,499	29,052,499	230,000
16p11.2	Variably expressive	16	29,652,499	30,202,499	550,000
17p13.3 *YWHAE*	Variably expressive	17	1,203,250	1,323,250	120,000
17q12	Variably expressive	17	34,815,887	36,225,887	1,410,000
17q23	Variably expressive	17	58,285,218	60,305,218	2,020,000
DiGeorge/VCFS	Variably expressive	22	19,020,000	20,290,000	1,270,000
22q11.2 distal	Variably expressive	22	21,910,000	23,670,000	1,760,000

Table S18. Genetic variants found in individuals with 16p12.1 deletion and family members (Excel file)

Table S19. Other hits identified from exome-sequencing data in 16p12.1 deletion cohort (Excel file)

Supplementary References

1. 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(14):1321-1331.

2. Iossifov I, O'Roak BJ, Sanders SJ, et al. The contribution of de novo coding mutations to autism spectrum disorder. *Nature.* 2014;515(7526):216-221.

3. Krumm N, Turner TN, Baker C, et al. Excess of rare, inherited truncating mutations in autism. *Nature genetics.* 2015;47(6):582-588.

4. 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(3):275-283.

5. 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(3):310-315.

6. 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(11):1152-1160.

7. 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.

8. 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(4):515-526.

9. 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(5):491-498.

10. McKenna A, Hanna M, Banks E, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome research.* 2010;20(9):1297-1303.

11. 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 10 11-33.

12. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics.* 2010;26(5):589-595.

13. 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(8):e1003709.

14. Schwarz JM, Rodelsperger C, Schuelke M, Seelow D. MutationTaster evaluates disease-causing potential of sequence alterations. *Nature methods.* 2010;7(8):575-576.

15. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic acids research.* 2010;38(16):e164.

16. Robinson JT, Thorvaldsdóttir H, Winckler W, et al. Integrative genomics viewer. *Nature biotechnology.* 2011;29(1):24-26.

17. Farrell MS, Werge T, Sklar P, et al. Evaluating historical candidate genes for schizophrenia. *Molecular psychiatry.* 2015;20(5):555-562.

18. Fromer M, Pocklington AJ, Kavanagh DH, et al. De novo mutations in schizophrenia implicate synaptic networks. *Nature.* 2014;506(7487):179-184.

19. Gilissen C, Hehir-Kwa JY, Thung DT, et al. Genome sequencing identifies major causes of severe intellectual disability. *Nature.* 2014;511(7509):344-347.

20. Girard SL, Gauthier J, Noreau A, et al. Increased exonic de novo mutation rate in individuals with schizophrenia. *Nature genetics.* 2011;43(9):860-863.

21. Greenwood TA, Light GA, Swerdlow NR, Radant AD, Braff DL. Association analysis of 94 candidate genes and schizophrenia-related endophenotypes. *PloS one.* 2012;7(1):e29630.

22. Grozeva D, Carss K, Spasic-Boskovic O, et al. Targeted Next-Generation Sequencing Analysis of 1,000 Individuals with Intellectual Disability. *Human mutation.* 2015;36(12):1197-1204.

23. 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(41):E5600-5607.

24. Iossifov I, Ronemus M, Levy D, et al. De novo gene disruptions in children on the autistic spectrum. *Neuron.* 2012;74(2):285-299.

25. Kaufman L, Ayub M, Vincent JB. The genetic basis of non-syndromic intellectual disability: a review. *Journal of neurodevelopmental disorders.* 2010;2(4):182-209.

26. Kirov G, Pocklington AJ, Holmans P, et al. De novo CNV analysis implicates specific abnormalities of postsynaptic signalling complexes in the pathogenesis of schizophrenia. *Molecular psychiatry.* 2012;17(2):142-153.

27. Levy D, Ronemus M, Yamrom B, et al. Rare de novo and transmitted copy-number variation in autistic spectrum disorders. *Neuron.* 2011;70(5):886-897.

28. Marshall CR, Noor A, Vincent JB, et al. Structural variation of chromosomes in autism spectrum disorder. *The American Journal of Human Genetics.* 2008;82(2):477-488.

29. Neale BM, Kou Y, Liu L, et al. Patterns and rates of exonic de novo mutations in autism spectrum disorders. *Nature.* 2012;485(7397):242-245.

30. 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(7397):246-250.

31. Purcell SM, Moran JL, Fromer M, et al. A polygenic burden of rare disruptive mutations in schizophrenia. *Nature.* 2014;506(7487):185-190.

32. 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(9854):1674-1682.

33. Ripke S, O'Dushlaine C, Chambert K, et al. Genome-wide association analysis identifies 13 new risk loci for schizophrenia. *Nat Genet.* 2013;45(10):1150-1159.

34. 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(12):1365-1369.

35. 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 research.* 2007;17(11):1665-1674.

36. Kramer A, Green J, Pollard J, Jr., Tugendreich S. Causal analysis approaches in Ingenuity Pathway Analysis. *Bioinformatics.* 2014;30(4):523-530.

37. Consortium GT. The Genotype-Tissue Expression (GTEx) project. *Nat Genet.* 2013;45(6):580-585.

38. Mi H, Muruganujan A, Casagrande JT, Thomas PD. Large-scale gene function analysis with the PANTHER classification system. *Nature protocols.* 2013;8(8):1551-1566.

39. 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(11):e13984.

40. Shannon P, Markiel A, Ozier O, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome research.* 2003;13(11):2498-2504.

41. 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(3):203-209.

42. Schmunk G, Nguyen RL, Ferguson DL, Kumar K, Parker I, Gargus JJ. High-throughput screen detects calcium signaling dysfunction in typical sporadic autism spectrum disorder. *Scientific reports.* 2017;7:40740.

43. Kwan V, Unda BK, Singh KK. Wnt signaling networks in autism spectrum disorder and intellectual disability. *Journal of neurodevelopmental disorders.* 2016;8:45.

44. Okerlund ND, Cheyette BN. Synaptic Wnt signaling-a contributor to major psychiatric disorders? *Journal of neurodevelopmental disorders.* 2011;3(2):162-174.