

This is an Open Access document downloaded from ORCA, Cardiff University's institutional repository:<https://orca.cardiff.ac.uk/id/eprint/175089/>

This is the author's version of a work that was submitted to / accepted for publication.

Citation for final published version:

Ali, Nabila M.H, Chawner, Samuel, Kushan-Wells, Leila, Bearden, Carrie, Mulle, Jennifer Gladys, Pollak, Rebecca, Gur, Raquel, Ghung, Wendy, Owen, Michael and Van Den Bree, Marianne 2024. Title comparison of autism domains across thirty rare variant genotypes. EBioMedicine

Publishers page:

Please note:

Changes made as a result of publishing processes such as copy-editing, formatting and page numbers may not be reflected in this version. For the definitive version of this publication, please refer to the published source. You are advised to consult the publisher's version if you wish to cite this paper.

This version is being made available in accordance with publisher policies. See <http://orca.cf.ac.uk/policies.html> for usage policies. Copyright and moral rights for publications made available in ORCA are retained by the copyright holders.



Title Comparison of Autism Domains Across Thirty Rare Variant Genotypes

Nabila M.H. Ali PhD ¹, Samuel J.R.A. Chawner PhD¹, Leila Kushan-Wells MSc ², Carrie E. Bearden PhD ², Jennifer Gladys Mulle PhD ^{3,4}, Rebecca M Pollak PhD ^{3,4}, Raquel E. Gur MD PhD ⁵, Wendy K. Chung MD PhD ⁶, IMAGINE ID consortium, Michael J. Owen FRCPsych ¹, Marianne B.M. van den Bree PhD¹

¹Centre for Neuropsychiatric Genetics and Genomics, Division of Psychological Medicine and Clinical Neurosciences, Cardiff University.

²Departments of Psychiatry and Behavioral Sciences and Psychology, Semel Institute for Neuroscience and Human Behavior, University of California Los Angeles.

³Department of Psychiatry, Robert Wood Johnson School of Medicine, Rutgers University

⁴Center for Advanced Biotechnology and Medicine, Rutgers University.

⁵Department of Psychiatry, Neurodevelopment & Psychosis Section, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA.

⁶Department of Pediatrics, Boston Children's Hospital and Harvard Medical School, Boston, MA, USA.

Word count: 6004

Structured Summary

Background

A number of Neurodevelopmental risk Copy Number Variants (ND-CNVs) and Single Gene Variants (SGVs) are strongly linked to elevated likelihood of autism. However, few studies have examined the impact on autism phenotypes across a wide range of rare variant genotypes.

Methods

This study compared Social Communication Questionnaire (SCQ) scores (total and subdomains: social, communication, repetitive behaviour) in 1,314 young people with one of thirty rare variant genotypes (15 ND-CNVs; n=1,005, 9.2 ± 3.5 years and 15 SGVs; n=309, 8.3 ± 4.0 years). Comparisons were also conducted with young people without known genetic conditions (controls; n=460, 10.6 ± 3.4 years) and with idiopathic autism (n=480, 8.6 ± 3.2 years).

Findings

The prevalence of indicative autism (SCQ ≥ 22) was higher in those with a rare variant genotype compared to controls (32% vs 2%; OR = 43.1, CI = 6.6 – 282.2, p < 0.001) and in those with SGVs compared to ND-CNVs (53% vs 25%; OR = 4.00, CI = 2.2 – 7.3, p = 0.002). The prevalence of indicative autism varied considerably across the 30 rare variant genotypes (range 10 – 85%). SGVs were associated with greater impairment in total, social, communication and repetitive behaviour subdomains than ND-CNVs. However, genotype explained limited variation in these scores (η^2 between 11.8–21.4%), indicating more convergence than divergence in autism phenotype across rare variant genotypes. Comparisons with young people with idiopathic autism indicated no differences compared to those with ND-CNVs, whereas those with SGVs showed greater communication and less repetitive behaviour.

Interpretation

The likelihood of autism was higher across all rare variant genotypes, with individuals with SGVs showing higher prevalence and greater impairment compared to those with ND-CNVs. Despite subdomain-specific patterns, there was no strong evidence for specific genotype-phenotype associations. This suggests that rare variant genotypes alone may have limited predictive value for autism phenotypes and that other factors like polygenic risk and the environment are likely to play a role. Further research is needed to understand these influences to improve risk prediction and inform genetic counselling and interventions.

Funding

This work was funded by the Tackling Multimorbidity at Scale Strategic Priorities Fund programme (MR/W014416/1) (van den Bree) delivered by the Medical Research Council and the National Institute for Health Research in partnership with the Economic and Social Research Council and in collaboration with the Engineering and Physical Sciences Research Council. NIMH U01 MH119738-01 (van den Bree), IMAGINE study ((Medical Research Council UK: MR/T033045/1; MR/N022572/1 and MR/L011166/1) (van den Bree) and Medical Research Council UK Centre Grant (MR/L010305/1) (Owen). SJRAC is funded by a Medical Research Foundation Fellowship (**MRF-058-0015-F-CHAW**). We would also like to acknowledge NIH 1R01MH110701-01A1 (PI Mulle), U01MH119736 (CEB), R21MH116473 (CEB), and R01MH085953 (CEB), and the Simons Foundation (SFARI Explorer Award to CEB).

Keywords

Autism, rare genetic variants, copy number variants (CNVs), single gene variants (SGVs).

Research in context

Evidence before this study

Autism is characterised by impairments in social communication, restricted interests, and repetitive behaviour. It is commonly associated with other neurodevelopmental disorders, such as attention deficit hyperactivity disorder (ADHD) and learning disabilities. Genetic factors play a considerable role in autism, with rare Neurodevelopmental risk Copy Number Variants (ND-CNVs) and Single Gene Variants (SGVs) exerting larger effects than common genetic variants. We searched titles and abstracts in PubMed published until September 2024 for studies in English that included the terms “autism”, “copy number variants”, and “single gene variants”, as well as each of the 30 rare variant genotypes that are part of our study. We found no clear evidence addressing the question of whether specific rare genetic variants were associated with specific autism subdomains. It is also not clear from the current literature whether autism in individuals with rare variant genotypes differs from idiopathic autism (i.e. with no known genetic origin).

The added value of this study

The majority of studies examining genotype-phenotype relationships within autism to date have included small samples and limited numbers of rare variant genotypes. Our international collaboration brought together a relatively large sample of young people (n=2,243) comprising 1,314 with one of 30 rare variant genotypes, 480 with idiopathic autism and 460 neurotypical individuals without known genetic conditions (controls). Indicative autism prevalence and subdomain scores were compared across rare variant genotypes and differences with individuals with idiopathic autism were investigated using the social communication questionnaire (SCQ).

Having a rare variant genotype was associated with a 43-fold increased likelihood of indicative autism. Individuals with SGVs had a higher prevalence of indicative autism than those with ND-CNVs. The prevalence of indicative autism varied considerably between rare variant genotypes (ranging from 10% to 85%). Similarly, considerable differences were found in autism total and subdomain scores across rare variant genotypes. Comparison of autism subdomains

between those with rare variant genotypes and those with idiopathic autism indicated comparable profiles for those with ND-CNVs, whereas those with SGVs showed greater impairment in the communication subdomain and less impairment in the repetitive behaviour subdomain. An individual's specific variant genotype explained between 11.8-21.4% of the variation in indicative autism total and subdomain scores.

Implications of all the available evidence

Rare variant genotypes are strongly associated with autism, with large variations in indicative autism prevalence and subdomain scores. Although rare variant genotypes showed subdomain-specific patterns, we did not find strong evidence of specific genotype-phenotype associations. Rather, our findings indicate a greater degree of convergence than divergence in autism prevalence and subdomain scores across rare variant genotypes. This suggests that rare variant genotypes alone may have limited predictive value for autism phenotypes and that other factors like polygenic risk and aspects of an individual's environment will also need to be taken into account to better understand the development of autism. Further study of the role of these factors in the development of autism in young people with rare variants will be important for the delivery of improved counselling and intervention services.

Introduction

Individuals with autism experience persistent impairment in social interaction and social communication, as well as an increase in restricted interests and repetitive behaviour¹. The genetic architecture underlying autism is complex, involving hundreds of common variants of small effect size^{2,3} and a range of rare Copy Number Variants (CNVs) and Single Gene Variants (SGVs) of relatively large effect size⁴⁻⁷. CNVs involve deletions or duplications of chromosomal segments larger than 1000 base pairs and are present in ~5-10% of individuals with autism⁸⁻¹⁰. Recently, the term Neurodevelopmental risk CNVs (ND-CNVs) has been used to refer to recurrent reciprocal CNVs associated with a high risk of neurodevelopmental disorder¹¹⁻¹³. Rare SGVs, disrupting specific genes, also contribute to autism aetiology and are estimated to be present in around ~10-30% of individuals with autism^{10,14,15}. The phenotypic presentation of individuals with rare variant genotypes is complex and involves other psychiatric conditions^{16,17} as well as cognitive impairment^{8,18-20}.

Whilst the number of rare variants reported to be associated with autism continues to grow^{4-6,21}, understanding of the phenotypic presentation is lacking. The extent to which different genotypes are associated with different aspects of the autism phenotype remains unclear, and predicting specific profiles in individuals with autism can be complicated by the frequently multimorbid presentation as well as incomplete penetrance and pleiotropy^{10,11,16,18}. These issues currently make using genetic information in the clinical setting highly challenging²². Although several studies have described the autism phenotype in individuals with rare variant genotypes²³⁻²⁵, there is a sparsity of studies examining the extent to which genetic heterogeneity underpins phenotypic heterogeneity^{24,26,27}.

Several studies have described the phenotypic profiles associated with specific rare variant genotypes, but these reports are limited by small sample sizes and the small number of variants investigated, e.g.,^{23,25-29}. Only a few studies to date^{24,25,28,30} have examined whether autism profiles differ between individuals with rare variant genotypes and those with idiopathic autism (i.e., autism of unknown genetic origin). Investigating these issues will elucidate whether individuals with autism with rare variant genotypes would benefit from different types of support compared

to those with idiopathic autism and inform care strategies. Previous work²⁴ recently compared individuals with deletions and duplications of 16p11.2 and 22q11.2 using the semi-structured research diagnostic Autism Diagnostic Interview (ADI-R)³¹. This work found subtle differences in autism profiles between the genotypes but much more substantial phenotypic variation within each genotype²⁴. Furthermore, over half (54%) of those who did not meet the diagnostic criteria for autism still exhibited clinically significant symptoms²⁴. These findings, however, focussed on only two genomic loci, and the extent to which the findings apply to a broader range of ND-CNVs as well as SGVs remains to be determined.

The Social Communication Questionnaire (SCQ) is a widely utilized tool to assess autism-related presentations in clinical practice and research³²⁻³⁴. It was developed based on the ADI-R³¹ and has been shown to have high specificity (80%)³⁵ and an acceptable diagnostic accuracy (area under the curve (AUC) = 0.88)³⁴, making it a cost-effective option for autism screening in large cohorts. A score of ≥ 22 on the SCQ indicates a need for a clinical evaluation of autism²⁹⁻³. However, it is important to note that the SCQ does not provide a definitive diagnosis of autism. In this work, we will refer to individuals who screened positive for autism on the SCQ as having “indicative autism” and to scores on the three sub-domains (social, communication and repetitive behaviour) as “autism subdomain scores”.

Our study aims to investigate differences in indicative autism scores across a range of rare variant genotypes. To address this question, we have assembled a sample of 1,314 young people with one of 30 rare variants, including 15 ND-CNVs (comprising losses and gains of various-sized segments across eight different chromosomes), as well as 15 SGVs. These 30 rare variant genotypes were selected because of a strong documented association with autism⁴⁻⁷. SCQ scores of these young people were compared to those of similar-aged neurotypical individuals without a known genetic condition (controls; n=460) as well as with individuals with idiopathic autism (n=480). We aimed to answer the following research questions:

- To what extent do autism prevalence and subdomain scores differ across rare variant genotypes, and what proportion of variation is explained by genotype?

- Do the autism subdomain scores of individuals with rare variant genotypes differ from those with idiopathic autism?

Methods

Study cohorts

Individuals with rare variant genotypes

1,314 individuals with either one of 15 ND-CNVs (n=1,005, mean age 9.2 ± 3.5 years, 60% male) or 15 SGVs (n=309, mean age 8.3 ± 4.0 years, 50% male) were identified across four different sites (Table 1). The ND-CNV cohort consisted of 1,005 young people with one of 15 ND-CNV across 8 chromosomal regions (9.2 ± 3.5 years, 60% male), which prior studies have associated with increased autism liability^{26,30,36,37} (Table 2). The SGVs cohort comprised 309 individuals (8.3 ± 4.0 years, 50% males) with variants in one of 15 genes. These variants included single nucleotide variants (de novo loss-of-function (LoF) variants, nonsense/frameshift variants) and deletions or duplications within genes known to be associated with autism³⁸⁻⁴⁰ (Table 2).

At Cardiff University, data was collected as part of the Cardiff University ECHO (Experiences of People With Copy Number Variants) <https://www.cardiff.ac.uk/centre-neuropsychiatric-genetics-genomics/research/themes/developmental-psychiatry/copy-number-variant-research-group> and the IMAGINE-ID Intellectual Disability and Mental Health: Assessing Genomic Impact on Neurodevelopment <https://www.ucl.ac.uk/child-health/research/population-policy-and-practice-research-and-teaching-department/cenb-clinical-29> studies (n=493). These two studies contributed data on 15 ND-CNVs (Supplementary Table 1).

The University of California Los Angeles (UCLA) contributed data on individuals with 22q11.2 deletion or duplication (n=79) (<https://www.semel.ucla.edu/bearden-lab>), and the 3q29 Project at Rutgers University on individuals with 3q29 deletion or duplication (n=69) (<https://sites.rutgers.edu/mulle/>). Finally, Simons Searchlight project

contributed data on individuals with 15 SGVs (n=309) and 5 CNVs ((1q21 deletion and duplication (n=64), 16p11.2 deletion (n=207) and 16p11.2 duplication (n=93)) (<https://www.sfari.org/resource/simons-searchlight/>). SGVs included single nucleotide variants (de novo loss-of-function (LoF) variants, nonsense/frameshift variants) and deletions or duplications within genes known to be associated with autism (Supplementary Table 1).

Individuals with idiopathic autism

The idiopathic autism cohort was recruited through the Simons Foundation Powering Autism Research for Knowledge (SPARK) project (<https://www.sfari.org/resource/spark/>). The SPARK registry comprises a cohort of approximately 99,000 children and adults with a clinical diagnosis of autism ⁴¹. For the current paper, we first excluded adults and individuals with incomplete/missing items on the SCQ and incomplete IQ data. Finally, we excluded all individuals with known rare pathogenic CNVs/ SGVs. We did not include this latter excluded group in our rare variant cohort because recruitment differences could have biased our findings. The idiopathic autism cohort included in this study thus comprised 480 young people (8.6 ± 3.2 years (81% male)) who were comparable to the rare variant genotypes cohort in age and sex (Table 1).

Neurotypical control participants

All sites that contributed individuals with rare variant genotypes also recruited neurotypical participants, resulting in a combined sample of 460 controls (10.6 ± 3.4 years, 53% male) (Table 1). These were either siblings of individuals with rare variant genotypes (familial controls, n=329, age (8.3 ± 4.0 years, 51% males)) or unrelated children (community-based controls, n=138, age (11.6 ± 3.5 years, 56% males)) ^{26,38,42,43}. The absence of neurodevelopmental risk variants in the siblings was confirmed through medical records and/or genotyping in the laboratories of the contributing sites. This information was, however, not available for the community-based controls. Both groups were comparable to the rare variant genotype cohorts in age and sex and were assessed with the same measures. Indicative autism prevalence (2% in familial controls and 1% in community-based controls) was comparable between the two control groups (OR=1.6, CI=0.36-8.1, p=0.736 (mixed effects logistic regression)) as were autism subdomain

scores (familial control mean total SCQ score 3.7 ± 4.1 , community-based controls mean total SCQ score 4.3 ± 5.6 , $p=0.910$). We, therefore, combined these two groups into one control cohort.

Phenotype assessments

Indicative autism prevalence and subdomain scores

In this study, all primary caregivers of participants - including those with rare variant genotypes, idiopathic autism, and controls - were requested to complete the SCQ to screen for autism. The SCQ consists of 40 yes/no questions that are scored based on the presence or absence of autism features³². The social subdomain score assesses the child's ability to interact socially, including their facial expressions, play, and ability to form friendships. This subdomain contributes 20 points to the total score. The communication subdomain score evaluates the child's use of language and nonverbal communication and contributes 11 points to the total score. The repetitive behaviour subdomain score assesses whether the child has restricted interests and engages in repetitive behaviour and contributes 8 points to the total score. These scores are commonly referred to as "autism subdomain scores"³².

The SCQ distinguishes between verbal and nonverbal children, asking, "Is she/he now able to talk using short phrases or sentences?". There are 7 questions that are only applicable to verbal children. Therefore, the total SCQ score ranges from 0-39 for verbal children, whereas for nonverbal children, it ranges from 0-32. Numbers and percentages of non-verbal participants were: ND-CNV ($n=80$ (8%)); SGV ($n= 176$ (57%)); idiopathic autism $n= 48$ (10%). We used a previously published method³² to adjust the total and communication subdomain scores for non-verbal participants to take into account the 7 missing items. We also performed a sensitivity analysis to assess how excluding nonverbal participants affected our findings.

A higher SCQ score usually indicates an increased likelihood of autism. We utilized the established cut-off of ≥ 22 as a positive indicator for autism screening³². This threshold effectively predicts autism, particularly in clinical populations^{44,45}, and serves as a clinical reference for further evaluation using the ADI-R. It's worth noting that prior research frequently used a cut-off of ≥ 15 to indicate potential autism spectrum traits^{26,32}. The SCQ has two versions:

lifetime and current ⁴⁶. The lifetime version queries observed behaviour across the child's lifespan, whereas the current version focuses on symptoms during the last three months. Although the two versions are functionally and psychometrically similar, the current version is used mainly for follow-up and evaluation of any interventions ³⁴. The lifetime version was administered in three cohorts (Cardiff, Simons Searchlight and SPARK; individuals with rare variant genotypes n=857, controls n= 374, idiopathic autism n=469) and the current version in two (the 3q29 Project and UCLA; individuals with rare variants n=148, controls n=86). The two versions were combined in analyses, and we conducted a sensitivity analysis to evaluate whether the findings were influenced by which version was used.

Cognitive profile

IQ assessments were conducted for individuals with ND-CNVs and controls from Cardiff University, as well as the Simons Searchlight and UCLA cohorts, but not the 3q29 Project cohort. IQ data was not available for individuals with SGVs. The sites that contributed IQ data used either the Wechsler Abbreviated Scale of Intelligence (WASI) ⁴⁷ or the Wechsler Intelligence Scale for Children (WISC-V)⁴⁸. Full-scale IQ (FSIQ), Performance IQ (PIQ), and Verbal IQ (VIQ) scores were derived from these scales ⁴⁹. The SPARK cohort obtained data on IQ for children with idiopathic autism from health records.

Other conditions associated with autism and medication use

Information on other conditions associated with autism was available for a subset of individuals with rare variant genotypes. These included attention deficit hyperactivity disorder (ADHD), seizures, oppositional defiant disorder, conduct disorder, depression, and tic disorder, as well as sleep problems as derived from semi-structured psychiatric interviews (Child & Adolescent Psychiatric Assessment (CAPA) at Cardiff University ²⁶) or clinical notes, and/or other primary carer-reported measures at the other sites ^{36,38,42,43,50} (Supplementary Table 2 for details).

Data on medication use (antipsychotics, anticonvulsants, mood stabilisers or ADHD treatment) (Supplementary Table 2), socioeconomic status (income and education) and ethnicity were obtained for those with ND-CNVs, SGVs,

idiopathic autism and controls. Medication use, socioeconomic status and ethnicity were included as covariates in sensitivity analyses.

Statistical Analysis

Indicative autism prevalence across individuals with rare variant genotypes

Mixed effects logistic regression models⁵¹ were used to determine whether the prevalence of indicative autism (outcome) differed by rare variant genotype (predictor). These comparisons were conducted between 1) individuals with rare variant genotypes versus controls; 2) individuals with rare variant genotypes (either ND-CNVs or SGVs) versus controls; 3) individuals with ND-CNVs versus those with SGVs; and 4) each of the 30 rare variant genotypes versus controls.

In each model, age and sex (self-reported by study participants or, where applicable, their caregiver) were included as fixed effects. In line with previous studies, study site (Europe versus United States) and family status (accounting for that fact that a subset of participants with rare variant genotypes and controls came from the same family) were included as random effects^{24,26,52}. The outcome of each model represented the odds of having indicative autism in each group compared to the other group. For comparison of the 30 rare variant genotypes versus controls, post hoc comparisons were subsequently conducted to determine group contrasts. All p-values were corrected for multiple comparisons using Tukey's Honest Significant Difference (HSD) adjustment⁵³.

Autism total and subdomain scores across individuals with rare variant genotypes

We used mixed-effects linear regression models⁵¹ to compare autism total and subdomain scores (the outcomes) first between individuals with ND-CNVs versus those with SGVs and then across the 30 different variant genotypes. We included study site and family status as random effects and age and sex as fixed effects. Post-hoc contrasts were conducted to determine between-group contrast estimates with Tukey's HSD adjustment of p-values⁵³.

Variation in autism total and subdomain scores explained by rare variant genotypes

To examine the variation in autism total and domain scores between rare variant genotypes, we conducted analysis of covariance (ANCOVA). In each ANCOVA model, the outcome (SCQ total and each of the three subdomain scores) was predicted by rare variant genotype with age, sex, and study site as covariates. We determined the proportion of variance in these scores that is attributable to rare variant genotype (between-genotype variation) using eta-square (η^2)⁵⁴, as in previous studies²⁴. The within-genotype variation was then calculated as the variation remaining after accounting for between-genotype variation and variation attributed to covariates (age, sex, and study site)(the three sources of variation summing to 100).

Autism total and subdomain scores in individuals with rare variant genotypes who screen positive for autism and individuals with idiopathic autism

We conducted comparisons between individuals with rare variant genotypes who tested positive for autism (SCQ score ≥ 22): ND-CNVs with indicative autism (ND-CNVs_ia, n=238) and SGVs with indicative autism (SGVs_ia; n=120) compared to individuals with idiopathic autism (n=480). Mixed-effects linear regression models were conducted, as explained above. We also assessed whether specific rare variant genotypes yield different total and subdomain scores when compared to individuals with idiopathic autism. This analysis included rare variant genotype groups that had a minimum of 10 participants who scored positively for autism, specifically 14 rare variant genotypes (9 ND-CNVs_ia; n=199 and 5 SGVs_ia; n=70). Post-hoc contrasts were conducted to determine between-group contrast estimates with Tukey's HSD adjustment of p-values⁵³.

Sensitivity analyses

Ethnic background, socioeconomic status, other conditions associated with autism and medication use

Sensitivity analyses were performed by including ethnic background, socioeconomic status, relevant conditions, and medication use (Supplementary Table 2) as covariates in the models outlined earlier to determine if these factors influenced our findings.

Inclusion of IQ as a covariate

We accounted for FSIQ, VIQ, and PIQ individually in each model mentioned above. Since not all sites assessed IQ, this analysis was confined to a subsample of 701 individuals with rare variant genotypes (53.3%), 320 controls (69.6%), and 480 individuals with idiopathic autism (100%).

Analysis of data using the lifetime SCQ version only

The Cardiff, Simons Searchlight, and SPARK cohorts used the SCQ lifetime version (individuals with rare variants n=1166, controls n=374, idiopathic autism n=469), while the current version was used by the 3q29 project (individuals with rare variants n=69, controls n=41) and UCLA (individuals with rare variants n=79, controls n=45). To evaluate if the inclusion of both versions in our analysis impacted the findings, we ran our models excluding data collected with the SCQ current version (administered by the 3q29 project and UCLA). This analysis was based on 1166 individuals with rare variant genotypes (84.9% of the total sample) and 374 controls (81.3% of the total sample).

Exclusion of nonverbal individuals

The ND-CNV cohort included 80 (8%), the SGV cohort 176 (57%) and the idiopathic autism cohort 48 (10%) nonverbal individuals. To evaluate the impact of including nonverbal participants, we reran the models, excluding those who were nonverbal. This analysis was based on n=925 individuals with ND-CNVs (92% of the total cohort with ND-CNVs), n=133 (43%) individuals with SGVs, and n=421 (90%) individuals with idiopathic autism.

All analyses were performed using R version 4.4.1. An alpha level of 0.05 was used to calculate multiple testing thresholds.

Ethics

All procedures involving human subjects/patients were approved by the appropriate local ethics committees or institutional review boards^{26,36,41-43,50,55,56}. Recruitment and assessment protocols for the ECHO study were approved by The South-East Wales Research Ethics Committee (09/WSE04/22).⁵⁵, while those of the IMAGINE-ID²⁶ were approved by the NHS London Queen Square research ethics committee (14/LO/1069). The 3q29 project was

approved by Emory University's Institutional Review Board (IRB00064133) and Rutgers University's Institutional Review Board (Pro2021001708). The UCLA Institutional Review Board approved the UCLA Study⁵⁶. Further details on Simon's Searchlight and the SPARK cohorts' ethical approval can be found at the website <https://www.sfari.org/resource/simons-searchlight/>^{38,41}.

Before recruitment, written consent or assent was obtained from each participant and, where applicable, their caregiver. The presence of rare variant genotypes was confirmed through microarray analysis at the laboratories of participating clinical research sites and/ or collected via medical records.

Role of funders

None of the funders had any role in study design, data collection, data analyses, interpretation, or the writing of this manuscript.

Results

Indicative autism prevalence across individuals with rare variant genotypes

Individuals with rare variant genotypes had a higher prevalence of indicative autism than controls (32% vs 2%; OR= 43.1, CI= 6.6 – 282.2, $p < 0.001$ (mixed effects logistic regression)). Those with SGVs were more likely to have indicative autism compared to those with ND-CNVs (53% vs 25%; OR= 4.00, CI= 2.2 – 7.3, $p = 0.002$ (mixed effects logistic regression)). We also calculated these prevalences using a cut-off of ≥ 15 on the SCQ to allow for comparison with other studies that used the same threshold. As anticipated, the rates increase with the less strict cut-off; however, the pattern remains consistent (Supplementary Table 3).

The prevalence of indicative autism was variable across the 30 rare variant genotypes, being highest in individuals with *ADNP* variants (85%, OR in comparison to controls = 489.8, CI=80.08 – 2995.41, $p < 0.001$ (mixed effects logistic regression)) and lowest in individuals with 22q11.2 deletion (10%, OR in comparison to controls=4.2, CI=1.94 – 9.02, $p = 0.042$ (mixed effects logistic regression)) (Figure 1.a, Supplementary table 4). However, it should be noted that

wide and overlapping confidence intervals existed for these estimates (Figure 1.b). Comparing the prevalence of indicative autism across the 30 rare variant genotypes revealed several significant differences (34 out of 435), mostly showing lower prevalence in those with 22q11.2 deletions or duplications or 16p11.2 deletions or duplications compared to those with SGVs (Supplementary Table 5).

Autism total and subdomain scores across individuals with rare variant genotypes

Individuals with SGVs had higher SCQ total scores than those with ND-CNVs (group contrast estimates=6.9, CI=4.6 – 9.3, $p<0.001$ (mixed effects linear regression)), indicating greater social disability in these groups. They also showed greater impairment in the social (group contrast estimates=3.9, CI=2.6 – 5.3, $p<0.001$ (mixed effects linear regression)), repetitive behaviour (group contrast estimates=0.9, CI=0.3 – 1.4, $p=0.009$ (mixed effects linear regression)) and communication subdomains (group contrast estimates=1.9, CI=1.3 – 2.7, $p<0.001$ (mixed effects linear regression)) (Figure 2.a, Supplementary Table 6).

Comparisons of total and subdomain scores across rare variant genotypes showed that individuals with 22q11.2 deletion generally showed the least impairment, whereas the opposite was true for those with *ADNP* variants (Figure 2.b, Supplementary Table 7). Subdomain-specific patterns were also present; for example, individuals with *STXBP1* were relatively impaired in the communication subdomain, those with *SCN2A* variants in the social, and those with *MED13L* variants in the repetitive behaviour subdomain. On the other hand, individuals with 1q21 TAR duplication showed relative strength in the communication subdomain, those with *HNRNPH2* variants in the repetitive behaviour, and those with *SETBP1* variants in the social subdomain (Figure 2.b, Supplementary Table 7).

Variation in autism total and subdomain scores explained by rare variant genotypes

Rare variant genotypes contributed to variation in autism total as well as the three subdomain scores. However, the proportion of variance explained was relatively low (<22%). Indeed, the variation in autism total score within rare variant genotypes was substantially greater than between-genotype ($\eta^2=73.7\%$ vs $\eta^2=21.4\%$, $p<0.0001$ (analysis of covariance)). Similarly, the within-genotype variation for the three subdomain scores ($\eta^2=75\%$ for social subdomain,

$\eta^2=78.8\%$ for communication subdomain and $\eta^2=84.6\%$ for repetitive behaviour subdomain) was considerably greater than between-genotype variation ($\eta^2=17.9\%$ for social subdomain ($p<0.0001$ (analysis of covariance)), $\eta^2=19.4\%$ for communication subdomain ($p<0.0001$ (analysis of covariance)) and, $\eta^2=11.8\%$ for repetitive behaviour subdomain ($p<0.0001$ (analysis of covariance)) (Figure 2.c, Supplementary Table 8).

Autism total and subdomain scores in individuals with rare variant genotypes who screen positive for autism and individuals with idiopathic autism

Individuals with ND-CNVs_ia (n=252) showed similar total and subdomain scores to those with idiopathic autism (Figure 3, Supplementary Table 9). Individuals with SGVs_ia (n=163) also showed comparable total score to those with idiopathic autism. However, they showed greater impairment in the communication subdomain (group contrast estimates=0.46, CI=0.17–0.75, $p=0.012$ (mixed effects linear regression)) and less impairment in the repetitive behaviour subdomain (group contrast estimates=-1.18, CI=-1.51 – -0.88, $p<0.001$ (mixed effects linear regression)) (Figure 3, Supplementary Table 9).

Post hoc analysis comparing individuals with rare variants who scored positive for autism with those with idiopathic autism showed several differences (Supplementary Table 10). Individuals with *SCN2A* variants showed more social impairment (group contrast estimates=2.71, CI=1.59 – 3.83, $p<0.001$ (mixed effects linear regression)) but less impairment in the repetitive behaviour subdomain (group contrast estimates=-2.23, CI=-2.85 – -1.61, $p<0.001$ (mixed effects linear regression)). Those with *GRIN2B* variants and 22q11.2 deletion showed less impairment in the repetitive behaviour subdomain (group contrast estimates= -2.43, CI=-3.3 – 1.56, $p<0.001$; group contrast estimates= -1.37, CI=-2.09 – -0.64, $p=0.028$ (mixed effects linear regression)) respectively. Furthermore, individuals with 22q11.2 deletion showed less impairment in the communication subdomain compared to those with idiopathic autism (group contrast estimates=-1.48, CI=-2.18 – -0.78, $p=0.006$ (mixed effects linear regression)) (Supplementary Table 10).

Sensitivity analysis

Ethnic background, socioeconomic status, other conditions associated with autism and medication

All findings remained consistent after adjusting for ethnicity, socioeconomic status, medication usage, and other conditions (Supplementary Table 11).

Inclusion of IQ as a covariate

The inclusion of FSIQ, Verbal IQ, or PIQ as a covariate in the analysis did not change the findings comparing between ND-CNVs and controls or across ND-CNVs (Supplementary Table 12).

Analysis of data using the lifetime SCQ version only

The exclusion of individuals assessed with the SCQ current rather than the lifetime version also did not impact the results (Supplementary Table 13).

Exclusion of nonverbal individuals

After excluding nonverbal individuals, the results stayed consistent, except for the loss of difference in the repetitive subdomain between individuals with SGVs_ia and those with idiopathic autism (Supplementary Table 14).

Discussion

Elucidation of the contribution of different rare variant genotypes to variability in autism phenotype requires sufficient numbers of participants across a range of genotypes assessed with the same measures^{22,24,26,27}. Through international collaboration, this study allowed the comparison of autism subdomain scores in a cohort of 2,254 participants, including 1,314 young people with one of 30 distinct ND-CNVs or SGVs that have previously been robustly associated with autism, alongside 480 individuals with idiopathic autism and 460 neurotypical controls of comparable age.

The presence of a rare variant genotype was associated with a 43-fold increased likelihood of indicative autism compared to controls. Our findings highlighted differences in indicative autism prevalence and subdomain profiles

between rare variant genotypes. Specifically, individuals with SGVs had a higher prevalence of indicative autism and were more impaired in the social, communication and repetitive behaviour subdomains compared to those with ND-CNVs. Furthermore, individuals with SGVs who screened positive for autism (SGVs_ia) showed a higher burden of communication difficulties and a lower burden of repetitive behaviour compared to those with idiopathic autism. Genotype explained between 11.7% to 21.4% of the variation in autism total and subdomain scores, whereas the variation within the genotypes was substantially larger (between 73.7% and 84.6%). Although our findings indicate that individual rare variant genotypes vary in prevalence and subdomain scores, overall, there is greater evidence of convergence of phenotype across genotypes rather than divergence. It is important to emphasise, however, that considerable heterogeneity existed within the genotypes, highlighting that these rare variants are not fully penetrant for autism and that other factors will also be important for the prediction of phenotypic variability.

Many ND-CNVs and SGVs have now been implicated in autism liability^{4-6,17}. However, a limited number of studies to date have compared autism phenotypes across a range of variants^{23,25-29}. We observed considerable variation in indicative autism prevalence across ND-CNVs and SGVs, with the presence of SGVs being associated with a higher prevalence of indicative autism than ND-CNVs. Although there was considerable variation across SGV genotypes, 9 of the 10 most penetrant variants were SGVs (Figure 2. B). This might be expected since the specific SGVs we studied were selected based on their strong association with autism³⁹. The lowest prevalence of autism was found in individuals with 22q11.2 deletion (10%, OR=4.19; CI=1.94 – 9.02, p=0.042 (mixed effects logistic regression)), consistent with reports comparing autism prevalence in individuals with 22q11.2 to those with other ND-CNVs^{23,24,57}. These findings are similarly consistent with previous research that has reported that autism penetrance varies across genetic aetiologies^{23,24,49,58,59}.

Previous studies have reported qualitative and quantitative differences in autism phenotype in individuals with rare variant genotypes^{24,26-29}. Our findings indicate that individuals with SGVs were more impaired in terms of having higher autism domain scores than those with ND-CNVs. SGVs are known to impact cognitive function, and our

findings are consistent with recent research reporting individuals with SGVs are more likely to exhibit severe autism symptoms and suggest higher comorbidity with intellectual disability compared to those with ND-CNVs^{21,60,61}. Our findings stress the need for further detailed comparative studies encompassing individuals with both SGVs and ND-CNVs. This is a research area of growing importance as SGVs are increasingly diagnosed in clinical settings.

We observed distinct patterns across the 30 rare variant genotypes, as illustrated in Figure 2.B. Those with 22q11.2 deletion were generally least impaired across total and subdomain scores, while those with *ADNP* variants tended to be most impaired. Subdomain-specific patterns were also found, indicating relative strengths for those with 1q21 TAR duplication in the communication subdomain, those with 16p11.2 deletion in the social subdomain and those with 16p11.2 deletion and *HNRNPH2* variation in the repetitive behaviour subdomain. Relative weaknesses were found in the communication subdomain for those with *STXBP1* variation, the social subdomain for those with *SC2NA* variation, and the repetitive behaviour subdomain for those with 15q11.2 deletion and *ASXL3* variation. These findings underscore the importance of comprehensive phenotypic assessments of rare variant genotypes beyond categorical autism diagnosis to allow fine-grained phenotypic characterisation.

Earlier research compared autism profiles in individuals with deletions or duplications of 16p11.2 and 22q11.2 to those with idiopathic autism, reporting subtle differences^{24,30}. Our findings, including a considerably broader range of ND-CNVs, also indicate differences in autism subdomain scores of individuals with ND-CNVs who scored positive for autism (ND-CNVs_ia) compared to those with idiopathic autism, although these were generally not significant. We did, however, find that individuals with 22q11.2 deletion exhibited less impairment in the communication and repetitive behaviour subdomains, which is in line with previous findings^{24,26,30,62}. Regarding comparisons between individuals with SGVs_ia and those with idiopathic autism, we did, uncover differences which indicated relative weaknesses in communication and relative strengths in the repetitive behaviour subdomain in those in the SGVs_ia. Our findings also show that the social subdomain scores of individuals with the SGV *SC2NA* indicated greater

impairment compared to those with idiopathic autism. These findings may be partially explained by the large number of non-verbal individuals in the SGVs_ia group.

We found the within-genotype variation in autism domain scores to be greater (73.7- 84.6%) than the between-genotype variation (21.4% for the total score, 17.9% for the social, 19.4% for the communication and 11.8% for the repetitive behaviour subdomain scores). This aligns with and extends previous research comparing autism domains across a more limited number of rare variant genotypes (deletion and duplication at 22q11.2 and 16p11.2)²⁴. This goes against the notion that rare variants lead to distinct autism subtypes and generally provides greater evidence of phenotypic convergence rather than divergence across genotypes. The picture is, however, complex, and it should be noted that the between-genotype variation was significant and that the reported effect sizes can be considered large (e.g., eta-squared >.14) as defined by Cohen ⁶³ , indicating that genotype-specific patterns also exist.

The findings have implications for the clinical setting as well as for future research in this area. Clinically, a generalised approach to assessing autism and providing support to individuals with different rare variant genotypes may be useful. Awareness amongst clinicians that, as a group, individuals with a wide range of rare variants have an increased risk of autism will be helpful, as clinical awareness of each individual variant and the associated phenotypic manifestations is often limited ⁶⁴. Yet, our findings of variation in prevalence and subdomain scores across genotypes also indicate the need for genotype-specific support. We detected a sizeable within-genotype variability that has important implications for future research as well as, more down the line, clinically. It suggests that other factors that are not yet well understood are likely to be important in understanding the expressivity of autism phenotypes across rare variant genotypes. Several studies have indicated that common variants with small effect sizes (Odds ratios <1.1) ^{2,65} can contribute to up to 50% of autism genetic heterogeneity ^{66,67} . This needs to be investigated further in individuals with rare variant genotypes. Additionally, further studies are necessary to assess the likelihood of autism in individuals with rare variants, especially concerning the influence of environmental risk factors like complications during pregnancy ⁶⁸ or adverse childhood experiences ⁶⁹. Advances in our understanding of these

additional factors and how they impact phenotypic expression in similar or different ways across rare variant genotypes can ultimately lead to more tailored support and care.

In summary, this study detected considerable variation in autism prevalence and subdomain scores across individuals with rare variant genotypes. The greater within-genotype than between-genotype variation indicates a lack of evidence of a solid, distinctive genotype-phenotype correlation between the rare variant genotypes studied and autism. The subdomain scores of individuals with ND-CNVs who scored positive for autism were comparable to those with idiopathic autism. At the same time, those with SGVs who scored positive for autism showed relative weaknesses in communication and relative strengths in the repetitive behaviour subdomain. We recommend that future research assesses individuals with rare genetic variant genotypes, combining standardised clinical autism assessment with detailed and wide-ranging phenotyping (including other neurodevelopmental and mental health conditions, as well as exposure to environmental risk factors) to further increase understanding of autism liability across these genotypes. In addition, the role of common genetic variation (as captured by polygenic risk scores) in the variable expressivity of these rare variants should also be clarified further ⁷⁰. Finally, future research must also focus on assessing the role of other genetic factors, such as disruptive variants elsewhere in the genome, in the prevalence of autism and profile presentation ^{9,17,71}.

Limitations

This study evaluated autism-related symptomatology using the SCQ, which shows high specificity for autism at the indicated cut-off point of ≥ 22 and is highly suitable for phenotyping autism in large cohorts ^{26,44,72}. Nevertheless, it should be noted that a formal diagnosis of autism would require more in-depth clinical diagnostic assessment, such as the ADI-R ³¹, and the findings cannot be generalised to individuals with other rare variant genotypes. It is also important to note that the SCQ has been reported to have reduced accuracy and discriminating capacity in screening for autism in individuals with more profound ID ⁷³. The rare variant genotypes examined in this study have all been associated with intellectual disability, although with varying rates and severity. We conducted a sensitivity analysis

taking into account the impact of IQ and found our findings remained largely unchanged. However, information on IQ was available in only 701 (53.3%) individuals with rare variant genotypes. Some studies have suggested a higher SCQ cut-off for individuals with more profound ID, and this was one of the reasons we applied the cut-off of ≥ 22 recommended by the developers of the SCQ, rather than the one of ≥ 15 , which studies have used to screen for autism spectrum disorder^{44,45,72}.

Another factor to note is that we incorporated items from the SCQ current version for a subset of our cohort for which lifetime data were not available. Previous research has shown that the current version tends to be less accurate than the lifetime version, especially when assessing children with ID^{34,45,46}. We recognize that merging data from both versions may introduce confounding bias. Ideally, using the lifetime version throughout would have been preferable, but inclusion of participants from the two sites (The 3q29 project and UCLA cohort) that utilized the current version of the SCQ enabled us to increase the number of young people with rare variants (14.7% contribution) as well as controls (18.6% contribution). Without these sites, we would not have been able to include the very rare 3q29 deletion and duplication genotypes in our analysis, as the 3q29 project contributed 69 out of 74 of these participants (93.2%). Furthermore, inclusion of the current version of the SCQ also enabled this work to benefit from the contribution by UCLA of an enhanced number of young people with the 22q11.2 duplication (n=79 out of 212 participants (37.3%)). Importantly, it was encouraging that our sensitivity analyses indicated that removing individuals assessed with the current version did not change our findings.

Although, as far as we are aware, we had available the largest possible sample to date to conduct our analyses, it is important to note that for some rare variant genotypes, particularly the SGVs, we were still limited by sample size. This will have impacted our comparisons of individual rare variant genotypes in particular. With increased sample sizes, future studies may find greater evidence of differences in autism prevalence and subdomain scores between genotypes.

It is also important to consider the effect of ascertainment bias. Although we took this into account to a degree by covarying for study site, differences in referral for genetic testing between those with ND-CNVs and SGVs may still have impacted our comparisons of these individuals. It should be emphasised, though, that the considerable differences in prevalence and subdomain scores we found between the different rare variant genotypes (including within ND-CNVs and SGVs) are unlikely to be explained by ascertainment bias alone. Furthermore, all participants in this study had a genetic diagnosis. As not all individuals with rare variant genotypes will have phenotypic manifestations deemed sufficiently severe to warrant referral for genetic testing, this work may overestimate the true impact of autism on those with rare variant genotypes^{8,24,26}.

Conclusion

The presence of a rare genetic variant genotype was associated with a 43 times increased risk of autism, and autism prevalence and subdomain scores were found to vary considerably across genotypes. However, the variation in autism total and subdomain scores within genotypes was considerably greater than between genotypes, indicating lack of strong evidence that rare variant genotypes are linked with discrete autism behavioural phenotypes. This implies that the specific rare variant - on its own - has limited autism phenotype predictability and prognostic significance, warranting the need for better understanding of the role other factors (both genetic and environmental) to move towards tailored clinical management of young people with rare variants.

Contributors

MBMvdB acquired the funding for the project and supervised the project. NMHali and MBMvdB arranged and coordinated data collection. NMHali, SJRAC, MBMvdB and MJO were involved in the analysis and interpretation of the data. NMHali, SJRAC, MBMvdB and MJO drafted the manuscript, which the other authors revised. LKW and CEB contributed to data collection from UCLA and reviewed the manuscript. JGM and RMP contributed to data collection from Rutgers University and reviewed the manuscript. REG reviewed the manuscript. WKC contributed to data

collection from Simon's Searchlight and reviewed the manuscript. The IMAGINE-ID consortium contributed to the data collection from Cardiff University. NMHali, SJRAC and MBMvdB directly accessed and verified the underlying data. All authors read and approved the final version of the manuscript.

Data Sharing Statement

The data used in this study is available upon request. Please contact Prof. van den Bree (vandenbreemb@cardiff.ac.uk) for any data requests. Approved researchers can obtain the Simons Searchlight population dataset described in this study (<https://www.sfari.org/resource/simons-searchlight/>) by applying at <https://base.sfari.org>. Approved researchers can obtain the SPARK population dataset described in this study (<https://www.sfari.org/resource/spark/>) by applying at <https://base.sfari.org>. The full phenotypic IMAGINE dataset is available from the UK Data Archive under special license access (SN 8621). Requests for genotype or linked genotypic-phenotypic data can be made through the study's data access committee: <https://imagine-id.org/healthcare-professionals/datasharing/>.

Declaration of interests

MJO and MBMvdB report grants from Takeda Pharmaceuticals outside of the submitted work. MJO reports a grant from Akvivia Health outside of the submitted work. CEB reports participation in a One Mind Scientific Advisory Board that is unrelated to the content of this manuscript. All other authors declare no competing interests.

Acknowledgement

We are grateful to all the children and families who participated in this study, the UK National Health Service (NHS) Medical Genetic clinics, and charities including Unique, Max Appeal and 22Crew for their support. We thank the ECHO, IMAGINE-ID, 3q29 consortium and UCLA studies for contributing data. We thank the core laboratory team at the MRC Centre for Neuropsychiatric Genetics and Genomics at Cardiff University (Cardiff, UK) for sample management and genotyping. We are grateful to all the families at the participating Simons Searchlight sites and the

Simons Searchlight Consortium, formerly the Simons VIP Consortium. We appreciate obtaining access to data on individuals with CNVs and SGVs on SFARI Base. We are grateful to all of the families in SPARK, the SPARK clinical sites and SPARK staff. We appreciate obtaining access to recruit participants through the SPARK research match on SFARI Base. This study was facilitated by the Genes to Mental Health Consortium (<https://genes2mentalhealth.com/>).

Tables

Table 1 Demographic information and prevalence of indicative autism and being non-verbal across the cohorts

Cohorts	Sample size	Sex		Age		Indicative autism*		Non- verbal	
	N	N	% (male)	Mean	SD	N	%	N	%
ND-CNV cohort									
Cardiff University cohort	493	310	63	9.5	3.1	144	29	28	6
Simons Searchlight cohort	364	209	57	8.4	3.7	87	24	39	11
UCLA** cohort	79	40	51	11.6	3.5	3	4	1	1
The 3q29 Project cohort	69	40	58	9.1	3.7	18	26	12	17
Total	1005	599	60	9.2	3.5	252	25	80	8
SGV cohort									
Simons Searchlight cohort	309	153	50	8.3	4.0	163	53	176	57
Control cohort									
Cardiff University cohort	208	112	54	10.4	2.8	8	4	1	0
Simons Searchlight cohort	166	83	50	10.4	3.7	0	0	0	0
UCLA cohort	45	26	58	12.0	3.3	2	4	3	7
The 3q29 Project cohort	41	22	54	10.4	4.2	0	0	0	0
Total	460	243	53	10.6	3.4	10	2	4	1
Idiopathic autism cohort***									
SPARK cohort ****	480	389	81	8.6	3.2	480	100	59	12

ND-CNVs: neurodevelopmental risk copy number variants

SGVs: single gene variants

* Based on a cut-off score of ≥ 22 on the Social Communication Questionnaire (SCQ)

** UCLA University of California at Los Angeles.

*** Idiopathic autism = autism with unknown genetic origin

**** SPARK: Simons Foundation Powering Autism Research for Knowledge

Table 2 Demographic information and prevalence of indicative autism and being non-verbal for the rare variant genotypes

Rare variant genotypes	Sample size N	Sex		Age		Indicative autism*		Non-verbal	
		N	% Male	mean	SD	N	%	N	%
ND-CNVs **									
1q21.1 deletion	47	30	64	8.2	3.5	14	30	2	4.3
1q21.1 duplication	67	39	58	8.7	3.4	32	48	7	10.4
1q21.1 TAR duplication	14	8	57	8.5	2.7	3	21	2	14.3
NRXN1 deletion	19	16	84	8.4	2.7	8	42	2	10.5
3q29 deletion	61	39	64	9.2	3.8	14	23	10	16.4
3q29 duplication	13	5	38	8.5	3.3	6	46	3	23.3
Kleefstra syndrome	13	5	38	11.6	3.9	7	54	1	7.7
15q13.3 deletion	27	21	78	9.4	3.8	11	41	1	3.7
15q13.3 duplication	23	15	65	9.4	3.2	8	35	1	4.3
15q11.2 deletion	42	32	76	9.2	3.1	16	38	3	7.1
16p11.2 deletion	254	146	57	8.8	3.6	48	19	26	10.2
16p11.2 distal deletion	24	14	58	9.4	3.1	9	38	5	20.8
16p11.2 duplication	125	76	61	10.2	3.6	39	31	8	6.4
22q11.2 deletion	212	114	54	10.1	3.1	22	10	5	2.4
22q11.2 duplication	64	39	61	10.1	3.4	14	22	4	6.2
Total	1005	599	60	9.2	3.5	252	25	80	8
SGVs ***									
<i>ADNP</i>	13	8	62	9.5	4.2	11	85	7	53.8
<i>ASXL3</i>	20	10	50	8.9	3.9	14	70	17	85
<i>CTNNA1</i>	16	6	38	9.1	4.7	6	38	5	31.2
<i>DYRK1A</i>	13	9	69	8.5	4.8	11	85	8	61.5
<i>GRIN2B</i>	27	18	67	8.6	4.0	15	56	12	44.4
<i>HIVEP2</i>	11	8	73	7.3	2.8	3	27	4	36.4
<i>HNRNP2</i>	11	1	9	10.5	5.3	6	55	8	72.7
<i>MED13L</i>	11	10	91	7.9	4.3	6	55	8	72.7
<i>PACS1</i>	17	6	35	7.6	4.1	7	41	7	41.2
<i>PPP2R5D</i>	33	12	36	7.4	3.6	14	42	15	45.5
<i>SCN2A</i>	52	23	44	8.1	4.3	31	60	38	73.1
<i>SETBP1</i>	10	7	70	7.5	3.5	2	20	7	70
<i>SLC6A1</i>	27	15	56	8.1	3.8	7	26	6	22.2
<i>STXBP1</i>	26	9	35	8.6	4.1	19	73	23	88.5
<i>SYNGAP1</i>	22	11	50	7.6	3.3	11	50	11	50.0
Total	309	153	50	8.3	4.0	163	53	176	57

ND-CNVs: neurodevelopmental risk- copy number variations.

SGVs: single gene variants

* Based on a cut-off score of ≥ 22 on the Social Communication Questionnaire (SCQ)

**ND-CNV regions included in the study are:1q21.1 (critical region 145.3-147.3), 2p16.3 NRXN1 (critical region 50.1-51.2), 3q29 (critical region 192.5-198.0), 9q34(critical region 140.5-140.7), 15q11.2 (critical region 22.8-23.1), 15q13.3 (critical region 31.10- 32.4), 16p11.2 (critical region 28.8-30.2) and 22q11.2 (critical region 19.0-21.5).

*** SGVs included in the study are: ADNP variants (chromosome band 20q13.13); ASXL3 variants (chromosome band 18q12.1); CTNNA1 variants (chromosome band 3p22.1); DYRK1A variants (chromosome band 21q22.13); GRIN2B variants (chromosome band 12p13.1); HIVEP2 variants (chromosome band 6q24.2); HNRNP2 variants (chromosome band Xq22.1); MED13L variants (chromosome band 12q24.21); PACS1 variants (chromosome band 12q24.21); PPP2R5D variants (chromosome band 6p21.1); SCN2A variants (chromosome band 2q24.3); 13 SETBP1 variants (chromosome band 18q12.3); SLC6A1 variants (chromosome band 3p25.3); STXBP1 variants (chromosome band 9q34.11); and SYNGAP1 variants (chromosome band 6p21.32).

References

1. American Psychiatric A. Diagnostic and statistical manual of mental disorders: DSM-5™, 5th ed. Arlington, VA, US: American Psychiatric Publishing, Inc.; 2013.
2. Anney R, Klei L, Pinto D, et al. Individual common variants exert weak effects on the risk for autism spectrum disorders. *Hum Mol Genet* 2012; **21**(21): 4781-92.
3. Klei L, McClain LL, Mahjani B, et al. How rare and common risk variation jointly affect liability for autism spectrum disorder. *Mol Autism* 2021; **12**(1): 66.
4. Sebat J, Lakshmi B, Malhotra D, et al. Strong association of de novo copy number mutations with autism. *Science* 2007; **316**(5823): 445-9.
5. Havdahl A, Niarchou M, Starnawska A, Uddin M, van der Merwe C, Warriar V. Genetic contributions to autism spectrum disorder. *Psychological Medicine* 2021; **51**(13): 2260-73.
6. Marshall CR, Noor A, Vincent JB, et al. Structural variation of chromosomes in autism spectrum disorder. *American journal of human genetics* 2008; **82**(2): 477-88.
7. Vicari S, Napoli E, Cordeddu V, et al. Copy number variants in autism spectrum disorders. *Progress in Neuro-Psychopharmacology and Biological Psychiatry* 2019; **92**: 421-7.
8. Raznahan A, Won H, Glahn DC, Jacquemont S. Convergence and Divergence of Rare Genetic Disorders on Brain Phenotypes: A Review. *JAMA Psychiatry* 2022.
9. Gilman SR, Iossifov I, Levy D, Ronemus M, Wigler M, Vitkup D. Rare de novo variants associated with autism implicate a large functional network of genes involved in formation and function of synapses. *Neuron* 2011; **70**(5): 898-907.
10. Huguet G, Benabou M, Bourgeron T. The Genetics of Autism Spectrum Disorders. In: Sassone-Corsi P, Christen Y, eds. *A Time for Metabolism and Hormones*. Cham (CH): Springer Copyright 2016, The Author(s). 2016: 101-29.
11. Kirov G. CNVs in neuropsychiatric disorders. *Human Molecular Genetics* 2015; **24**(R1): R45-R9.
12. Takumi T, Tamada K. CNV biology in neurodevelopmental disorders. *Current Opinion in Neurobiology* 2018; **48**: 183-92.
13. Rees E, Kirov G. Copy number variation and neuropsychiatric illness. *Current Opinion in Genetics & Development* 2021; **68**: 57-63.
14. Huguet G, Ey E, Bourgeron T. The Genetic Landscapes of Autism Spectrum Disorders. *Annual Review of Genomics and Human Genetics* 2013; **14**(1): 191-213.
15. Iossifov I, Ronemus M, Levy D, et al. De novo gene disruptions in children on the autistic spectrum. *Neuron* 2012; **74**(2): 285-99.
16. Mollon J, Almasy L, Jacquemont S, Glahn DC. The contribution of copy number variants to psychiatric symptoms and cognitive ability. *Molecular Psychiatry* 2023; **28**(4): 1480-93.
17. Rylaarsdam L, Guemez-Gamboa A. Genetic Causes and Modifiers of Autism Spectrum Disorder. *Frontiers in cellular neuroscience* 2019; **13**: 385-.
18. Molloy CJ, Quigley C, McNicholas Á, Lisanti L, Gallagher L. A review of the cognitive impact of neurodevelopmental and neuropsychiatric associated copy number variants. *Transl Psychiatry* 2023; **13**(1): 116.

19. Jacquemont S, Huguet G, Klein M, et al. Genes To Mental Health (G2MH): A Framework to Map the Combined Effects of Rare and Common Variants on Dimensions of Cognition and Psychopathology. *American Journal of Psychiatry* 2022; **179**(3): 189-203.
20. Cunningham AC, Hall J, Owen MJ, van den Bree MBM. Coordination difficulties, IQ and psychopathology in children with high-risk copy number variants. *Psychol Med* 2021; **51**(2): 290-9.
21. Toma C. Genetic Variation across Phenotypic Severity of Autism. *Trends in Genetics* 2020; **36**(4): 228-31.
22. Binder EB. Genotype-Phenotype Predictions in Autism: Are We There Yet? *American Journal of Psychiatry* 2021; **178**(1): 11-2.
23. Richards C, Jones C, Groves L, Moss J, Oliver C. Prevalence of autism spectrum disorder phenomenology in genetic disorders: a systematic review and meta-analysis. *The Lancet Psychiatry* 2015; **2**(10): 909-16.
24. Chawner S, Doherty JL, Anney RJL, et al. A Genetics-First Approach to Dissecting the Heterogeneity of Autism: Phenotypic Comparison of Autism Risk Copy Number Variants. *Am J Psychiatry* 2021; **178**(1): 77-86.
25. Bozhilova N, Welham A, Adams D, et al. Profiles of autism characteristics in thirteen genetic syndromes: a machine learning approach. *Molecular Autism* 2023; **14**(1): 3.
26. Chawner SJRA, Owen MJ, Holmans P, et al. Genotype–phenotype associations in children with copy number variants associated with high neuropsychiatric risk in the UK (IMAGINE-ID): a case-control cohort study. *The Lancet Psychiatry* 2019; **6**(6): 493-505.
27. Bruining H, Eijkemans MJ, Kas MJ, Curran SR, Vorstman JA, Bolton PF. Behavioral signatures related to genetic disorders in autism. *Mol Autism* 2014; **5**(1): 11.
28. Bruining H, de Sonnevile L, Swaab H, et al. Dissecting the clinical heterogeneity of autism spectrum disorders through defined genotypes. *PLoS One* 2010; **5**(5): e10887.
29. Lee NR, Niu X, Zhang F, et al. Variegation of autism related traits across seven neurogenetic disorders. *Translational Psychiatry* 2022; **12**(1): 149.
30. Serur Y, Sofrin Frumer D, Daon K, et al. Psychiatric disorders and autism in young children with 22q11.2 deletion syndrome compared to children with idiopathic autism. *Eur Psychiatry* 2019; **55**: 116-21.
31. Lord C, Rutter M, Le Couteur A. Autism Diagnostic Interview-Revised: a revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *J Autism Dev Disord* 1994; **24**(5): 659-85.
32. Eaves LC, Wingert HD, Ho HH, Mickelson ECR. Screening for Autism Spectrum Disorders With the Social Communication Questionnaire. *Journal of Developmental & Behavioral Pediatrics* 2006; **27**(2).
33. Charman T, Baird G, Simonoff E, et al. Efficacy of three screening instruments in the identification of autistic-spectrum disorders. *British Journal of Psychiatry* 2007; **191**(6): 554-9.
34. Chesnut SR, Wei T, Barnard-Brak L, Richman DM. A meta-analysis of the social communication questionnaire: Screening for autism spectrum disorder. *Autism* 2017; **21**(8): 920-8.
35. Chandler S, Charman T, Baird G, et al. Validation of the social communication questionnaire in a population cohort of children with autism spectrum disorders. *J Am Acad Child Adolesc Psychiatry* 2007; **46**(10): 1324-32.
36. Pollak RM, Murphy MM, Epstein MP, et al. Neuropsychiatric phenotypes and a distinct constellation of ASD features in 3q29 deletion syndrome: results from the 3q29 registry. *Molecular Autism* 2019; **10**(1): 30.
37. Pollak RM, Zinsmeister MC, Murphy MM, Zwick ME, Mulle JG. New phenotypes associated with 3q29 duplication syndrome: Results from the 3q29 registry. *Am J Med Genet A* 2020; **182**(5): 1152-66.
38. The Simons Vip Consortium. Simons Variation in Individuals Project (Simons VIP): A Genetics-First Approach to Studying Autism Spectrum and Related Neurodevelopmental Disorders. *Neuron* 2012; **73**(6): 1063-7.

39. Arpi MNT, Simpson TI. SFARI genes and where to find them; modelling Autism Spectrum Disorder specific gene expression dysregulation with RNA-seq data. *Scientific Reports* 2022; **12**(1): 10158.
40. Abrahams BS, Arking DE, Campbell DB, et al. SFARI Gene 2.0: a community-driven knowledgebase for the autism spectrum disorders (ASDs). *Mol Autism* 2013; **4**(1): 36.
41. Feliciano P, Daniels AM, Green Snyder L, et al. SPARK: A US Cohort of 50,000 Families to Accelerate Autism Research. *Neuron* 2018; **97**(3): 488-93.
42. Pollak RM, Zinsmeister MC, Murphy MM, Zwick ME, the Emory 3q P, Mulle JG. New phenotypes associated with 3q29 duplication syndrome: Results from the 3q29 registry. *American Journal of Medical Genetics Part A* 2020; **182**(5): 1152-66.
43. Lin A, Vajdi A, Kushan-Wells L, et al. Reciprocal Copy Number Variations at 22q11.2 Produce Distinct and Convergent Neurobehavioral Impairments Relevant for Schizophrenia and Autism Spectrum Disorder. *Biological Psychiatry* 2020.
44. Allen CW, Silove N, Williams K, Hutchins P. Validity of the social communication questionnaire in assessing risk of autism in preschool children with developmental problems. *J Autism Dev Disord* 2007; **37**(7): 1272-8.
45. Hollocks MJ, Casson R, White C, Dobson J, Beazley P, Humphrey A. Brief Report: An Evaluation of the Social Communication Questionnaire as a Screening Tool for Autism Spectrum Disorder in Young People Referred to Child & Adolescent Mental Health Services. *J Autism Dev Disord* 2019; **49**(6): 2618-23.
46. Wei T, Chesnut SR, Barnard-Brak L, Richman D. Psychometric analysis of the Social Communication Questionnaire using an Item-Response Theory framework: Implications for the use of the lifetime and current forms. *Journal of Psychopathology and Behavioral Assessment* 2015; **37**(3): 469-80.
47. Axelrod BN. Validity of the Wechsler abbreviated scale of intelligence and other very short forms of estimating intellectual functioning. *Assessment* 2002; **9**(1): 17-23.
48. Saklofske DH, Breaux K, Beal AL, Raiford SE, Weiss LG. Chapter 9 - WISC-V and the Evolving Role of Intelligence Testing in the Assessment of Learning Disabilities. In: Weiss LG, Saklofske DH, Holdnack JA, Prifitera A, eds. WISC-V (Second Edition): *Academic Press*; 2019: 271-336.
49. 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**(1): 20-30.
50. Simons Variation in Individuals Project (Simons VIP): a genetics-first approach to studying autism spectrum and related neurodevelopmental disorders. *Neuron* 2012; **73**(6): 1063-7.
51. Meteyard L, Davies RAI. Best practice guidance for linear mixed-effects models in psychological science. *Journal of Memory and Language* 2020; **112**: 104092.
52. Niarchou M, Chawner S, Doherty JL, et al. Psychiatric disorders in children with 16p11.2 deletion and duplication. *Transl Psychiatry* 2019; **9**(1): 8.
53. Midway S, Robertson M, Flinn S, Kaller M. Comparing multiple comparisons: practical guidance for choosing the best multiple comparisons test. *PeerJ* 2020; **8**: e10387.
54. Richardson JTE. Eta squared and partial eta squared as measures of effect size in educational research. *Educational Research Review* 2011; **6**(2): 135-47.
55. Cunningham AC, Delpont S, Cumines W, et al. Developmental coordination disorder, psychopathology and IQ in 22q11.2 deletion syndrome. *Br J Psychiatry* 2018; **212**(1): 27-33.
56. Seitz-Holland J, Lyons M, Kushan L, et al. Opposing white matter microstructure abnormalities in 22q11.2 deletion and duplication carriers. *Transl Psychiatry* 2021; **11**(1): 580.

57. Wenger TL, Miller JS, DePolo LM, et al. 22q11.2 duplication syndrome: elevated rate of autism spectrum disorder and need for medical screening. *Mol Autism* 2016; **7**: 27.
58. Malhotra D, Sebat J. CNVs: harbingers of a rare variant revolution in psychiatric genetics. *Cell* 2012; **148**(6): 1223-41.
59. Rosenfeld JA, Coe BP, Eichler EE, Cuckle H, Shaffer LG. Estimates of penetrance for recurrent pathogenic copy-number variations. *Genet Med* 2013; **15**(6): 478-81.
60. Arnett AB, Rhoads CL, Hoekzema K, et al. The autism spectrum phenotype in ADNP syndrome. *Autism Res* 2018; **11**(9): 1300-10.
61. Dang T, Duan WY, Yu B, et al. Autism-associated Dyrk1a truncation mutants impair neuronal dendritic and spine growth and interfere with postnatal cortical development. *Molecular Psychiatry* 2018; **23**(3): 747-58.
62. Selten I, Boerma T, Everaert E, et al. Behaviors related to autism spectrum disorder in children with developmental language disorder and children with 22q11.2 deletion syndrome. *Autism & Developmental Language Impairments* 2023; **8**: 23969415231179844.
63. Cohen J. Statistical power analysis for the behavioral sciences: routledge; 1988.
64. Chawner SJRA, Watson CJ, Owen MJ. Clinical evaluation of patients with a neuropsychiatric risk copy number variant. *Current Opinion in Genetics & Development* 2021; **68**: 26-34.
65. Bray NJ, O'Donovan MC. The genetics of neuropsychiatric disorders. *Brain Neurosci Adv* 2019; **2**.
66. Chaste P, Roeder K, Devlin B. The Yin and Yang of Autism Genetics: How Rare De Novo and Common Variations Affect Liability. *Annual Review of Genomics and Human Genetics* 2017; **18**(1): 167-87.
67. Gaugler T, Klei L, Sanders SJ, et al. Most genetic risk for autism resides with common variation. *Nature genetics* 2014; **46**(8): 881-5.
68. Lu J, Wang Z, Liang Y, Yao P. Rethinking autism: the impact of maternal risk factors on autism development. *Am J Transl Res* 2022; **14**(2): 1136-45.
69. Baldwin JR, Sallis HM, Schoeler T, et al. A genetically informed Registered Report on adverse childhood experiences and mental health. *Nat Hum Behav* 2023; **7**(2): 269-90.
70. Niemi MEK, Martin HC, Rice DL, et al. Common genetic variants contribute to risk of rare severe neurodevelopmental disorders. *Nature* 2018; **562**(7726): 268-71.
71. Wiśniowiecka-Kowalnik B, Nowakowska BA. Genetics and epigenetics of autism spectrum disorder—current evidence in the field. *Journal of Applied Genetics* 2019; **60**(1): 37-47.
72. Charman T, Gotham K. Measurement Issues: Screening and diagnostic instruments for autism spectrum disorders – lessons from research and practise. *Child and Adolescent Mental Health* 2013; **18**(1): 52-63.
73. Sappok T, Diefenbacher A, Gaul I, Bölte S. Validity of the Social Communication Questionnaire in Adults With Intellectual Disabilities and Suspected Autism Spectrum Disorder. *American Journal on Intellectual and Developmental Disabilities* 2015; **120**(3): 203-14.

Figure Legends

Figures

Figure 1. Prevalence and Odds Ratios of indicative autism in individuals with rare variants genotypes

a. Prevalence of indicative autism in individuals with rare variant genotypes

The figure illustrates the percentages of individuals who passed the autism cut-off of SCQ ≥ 22 (indicative autism) in the 30 rare variant genotypes.

ND-CNVs = neurodevelopmental risk- copy number variants.

SGVs=single gene variants.

b. Odds ratios of having indicative autism in individuals with rare variant genotypes compared to controls

The figure illustrates the odds of having indicative autism in individuals with rare variant genotypes compared to controls. The ORs were derived from mixed-effects logistic regression models.

All p-values were significant after correction for multiple testing except for the comparisons of 1q21 TAR duplication and SETBP1 with the control sample.

ND-CNVs = neurodevelopmental risk- copy number variants.

SGVs=single gene variants.

Figure 2. Autism subdomain scores across individuals with rare variant genotypes

a. Autism total and subdomain scores in individuals with ND-CNVs compared to those with SGVs

This figure shows the mean total and subdomain scores in individuals with SGVs compared to those with ND-CNVs

ND-CNVs = neurodevelopmental risk- copy number variants.

SGVs=single gene variants

** $p < 0.01$, *** $p < 0.001$ (mixed effects linear regression)

b. Autism total and subdomain scores variation across individuals with rare variant genotypes

This heatmap plot was generated by standardising the total and subdomain scores of each rare variant genotype (compared to the mean score of all variants) using z scores. Darker colours indicate relatively higher scores in a domain (more autism symptoms are endorsed – greater impairment). It shows subdomain-specific patterns across the 30 rare variant genotypes.

c. Variation in autism total and subdomain scores explained by rare variant genotype

The figure illustrates the between-group and within-group variation in autism domain scores. The within-group variation is considerably larger than the between-group variation.

Between-genotype variation: variation explained by rare variant genotype, as measured η^2 (eta-squared).

Variation explained by covariates: age, sex and study site

Within-genotype variation: Variation remaining within each variant genotype after considering the between-genotype variation and variation explained by other covariates(see Methods for further explanation).

Figure 3 Autism subdomain scores in individuals with rare variant genotypes compared to those with idiopathic autism

This figure shows the mean total and subdomain scores in individuals with ND-CNVs and SGVs compared to those with idiopathic autism

ND-CNVs_ia: individuals with ND-CNVs who passed the cut-off of ≥ 22 for autism on the SCQ

SGVs_ia individuals with SGVs who passed the cut-off of ≥ 22 for autism on the SCQ

* $p < 0.05$, *** $p < 0.001$ (mixed effects linear regression).