A Genetics-First Approach to Dissecting the Heterogeneity of Autism: Phenotypic Comparison of Autism Risk Copy Number Variants

Samuel J.R.A. Chawner[1], PhD, Joanne L. Doherty[1], PhD MRCPsych, Richard Anney[1], PhD, Kevin M. Antshel[2], PhD, Carrie E. Bearden[3], PhD, Raphael Bernier[4], PhD, Wendy K. Chung[5], MD PhD, Caitlin C. Clements[6,7], MA, Sarah R. Curran[8], PhD MRCPsych, Goran Cuturilo[9], MD PhD, Ania M. Fiksinski[10,11], MSc, Louise Gallagher[12], MD PhD, Robin P. Goin-Kochel[13], Leila Kushan[3], MS, PhD, Raquel E. Gur[14,15,16], MD PhD, Ellen Hanson[17], PhD, Sebastien Jacquemont[18,19], MD, Wendy R. Kates[20], PhD, Anne M. Maillard[18,21], PhD, Donna M. McDonald-McGinn[22,23,24], MS LCGC, Marina Mihaljevic[25], MD PhD, Judith S Miller[6], PhD, Hayley Moss[1], BSc, Milica Pejovic-Milovancevic[26], MD PhD, Robert T. Schultz[6,24], PhD, LeeAnne Green-Snyder[27], PhD, Jacob A. Vorstman[10,28,29], MD PhD, Tara L. Wenger[7,30], MD PhD, IMAGINE-ID Consortium*, Jeremy Hall[1,31], PhD FRCPsych, Michael J. Owen[1,31], PhD FRCPsych, & Marianne B.M. van den Bree[1,31], PhD.

[1]MRC Centre for Neuropsychiatric Genetics and Genomics, Division of Psychological Medicine and Clinical Neurosciences, Cardiff University, Cardiff, UK
[2]Department of Psychology, Syracuse University, Syracuse, NY, USA
[3]Department of Psychiatry and Behavioral Sciences and Psychology, Semel Institute for Neuroscience and Human Behavior, University of California, Los Angeles, CA, USA
[4]Departments of Psychiatry and Behavioral Sciences, University of Washington, Seattle, WA, USA
[5]Columbia University Irving Medical Center Departments of Pediatrics and Medicine, New York, NY, USA
[6] Center for Autism Research, Children's Hospital of Philadelphia, 2716 South Street, Philadelphia, PA 19104, USA
[7]Department of Psychology, University of Pennsylvania, 3720 Walnut Street, Philadelphia, PA 19104, USA
[8]South West London and St George's Mental Health NHS Foundation Trust, London, UK
[9]University Children's Hospital, Belgrade, Serbia; Faculty of Medicine, University of Belgrade, Belgrade, Serbia
[10]Psychiatry, Brain Center Rudolf Magnus UMC Utrecht, Utrecht, Netherlands
[11] Clinical Genetics Research Program, Centre for Addiction and Mental Health, and The Dalglish Family 22q Clinic for 22q11.2 Deletion Syndrome, Toronto General Hospital, University Health Network, Toronto, Ontario, Canada
[12]Department of Psychiatry, Trinity College Dublin, Dublin, Ireland
[13]Department of Pediatrics, Baylor College of Medicine, Houston, TX, USA
[14]Lifespan Brain Institute, Children's Hospital of Philadelphia and Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA
[15]Department of Psychiatry, Neurodevelopment & Psychosis Section, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA
[16]Department of Child and Adolescent Psychiatry, Children's Hospital of Philadelphia, Philadelphia, PA, USA
[17]Developmental Medicine, Children's Hospital Boston/Harvard Medical School, Boston, MA, USA
[18]Departement of pediatrics, University of Montreal, Montréal, Qc, Canada.
[19]Centre de recherche, Centre Hospitalier Universitaire Sainte Justine, Montréal, Canada
[20]Department of Psychiatry and Behavioral Sciences, State University of New York at Upstate Medical University, Syracuse, NY, USA
[21] Service des Troubles du Spectre de l'Autisme et Apparentés, Lausanne University Hospital, Switzerland
[22]Division of Human Genetics and Molecular Biology, Children's Hospital of Philadelphia, 3401 Civic Center Boulevard, Philadelphia, PA, USA
[23]22q and You Center, Children's Hospital of Philadelphia, 3401 Civic Center Boulevard, Philadelphia, PA, USA
[24]Department of Pediatrics, Perelman School of Medicine of University of Pennsylvania, Philadelphia, PA, USA
[25] Clinic for Psychiatry, Clinical Center of Serbia; Faculty of Medicine, University of Belgrade, Belgrade, Serbia
[26]Institute of Mental Health, Belgrade, Serbia; Faculty of Medicine, University of Belgrade, Belgrade, Serbia
[27]Simons Foundation, 160 5th Avenue, 7th Floor, New York, NY, USA
[28]Program in Genetics and Genome Biology, SickKids Research Institute, Toronto, Ontario, Canada
[29]Department of Psychiatry, The Hospital for Sick Children, University of Toronto, Ontario, Canada
[30]Department of Pediatrics, Seattle Children's Hospital, M/S OB.9.5204800 Sand Point Way NE, Seattle, WA, USA
[31]Neuroscience and Mental Health Research Institute, Cardiff University, Cardiff, UK

*See IMAGINE-ID consortium member list

Abstract

Objective: Certain copy number variants (CNVs) greatly increase risk of autism. We conducted a genetics-first study to investigate whether heterogeneity in the clinical presentation of autism is underpinned by specific genotype-phenotype relationships.

Methods: This international study included 547 individuals (12.3 years (SD=4.2), 54% male) who were ascertained on the basis of having a genetic diagnosis of a rare CNV associated with high risk of autism (82 16p11.2 deletion carriers, 50 16p11.2 duplication carriers, 370 22q11.2 deletion carriers and 45 22q11.2 duplication carriers), as well as 2027 individuals (9.1 years (SD=4.9), 86% male) with autism of heterogeneous aetiology. The Autism Diagnostic Interview-Revised (ADI-R) and IQ testing were conducted.

Results: The four genetic variant groups differed in autism severity, autism subdomain profile as well as IQ profile. However, we found substantial variability in phenotypic outcome within individual genetic variant groups (74% to 97% of the variance depending on the trait), whereas variability between groups was low (1% to 21% depending on trait). We compared CNV carriers who met autism criteria, to individuals with heterogeneous autism, and a range of profile differences were identified. Using clinical cut-offs, we found that 54% of individuals with one of the 4 CNVs who did not meet full autism diagnostic criteria nonetheless had elevated levels of autistic traits.

Conclusion: Many CNV carriers do not meet full diagnostic criteria for autism, but nevertheless meet clinical cut-offs for autistic traits. Although we find profile differences between variants, there is considerable variability in clinical symptoms within the same variant.

Introduction

Autism is a behaviourally defined condition characterized by deficits in social interaction and communication, as well as the presence of restricted, repetitive behaviours and interests[1]. There is considerable heterogeneity in the clinical presentation of autism, in terms of symptom profile, cognitive function and developmental trajectories[2-5]. Studies of large genotyped cohorts of individuals with autism and typically developing controls have identified several chromosomal copy number variants (CNVs) (deletions and duplications >1 kilobase (kb)[6]) as genetic risk factors for autism[7-12], and have been demonstrated in clinical settings to have predictive value[13]. Although individually rare, collectively pathogenic CNVs are identified in 15% of patients with neurodevelopmental disability[14]. A number of researchers have advocated that the time is ripe for a reverse strategy based on a genetics-first rather than a phenotype-first approach, in order to better understand the clinical heterogeneity of autism[15-17].

Deletions and duplications at the 16p11.2 (600kb, break points 4 and 5 (BP4-BP5) critical region 29.6-30.2 Mb, build hg19) and 22q11.2 (3 Mb, break points A and D, critical region 19.0-21.5 Mb, build hg19) loci have been identified as risk factors for autism, both from phenotype-first studies that find these variants occur with greater frequency in cohorts of individuals with autism versus controls[7-9], and genetics-first studies which find that patients diagnosed with 16p11.2 and 22q11.2 CNVs in medical genetics clinics have an elevated frequency of autism diagnosis[18-25] relative to the frequency in the general population of 1% [26,27]. It is important to determine whether these variants confer risk for the same autism phenotype or whether the presentation differs by genotype. The former would indicate that genomic risk for autism has common phenotypic effects, whilst the latter would suggest that genetic heterogeneity underpins clinical heterogeneity. Within the autism field, there is a strong notion that the condition is dissociable by genetics[28,29], with some researchers using the term “autisms”[30]. Early evidence indicates that the 22q11.2 deletion and duplication may have unique autism profiles[25,31], however, the profiles of the two groups have not been directly compared within the same study, and hence the differences reported could be due to methodological

inconsistencies. For the 16p11.2 locus, it has been reported that duplication carriers with autism have lower IQ compared to deletion carriers with autism[18], however the autism profiles of the two groups have not been compared. It is also important to investigate the extent to which the autism profile of these variants differs from individuals without these variants who have autism (referred to as heterogeneous autism from here onwards).

Comprehensive clinical phenotyping of individuals with autism-risk genetic variants requires large integrated networks of researchers and clinicians using the same clinical instruments. The present study brings together patient data from several international genetics-first consortia of individuals with rare chromosomal conditions associated with high risk of autism. Individuals with deletions and duplications that span critical regions at the 22q11.2 and 16p11.2 loci were ascertained clinically via medical genetics clinics and patient organisations. We aimed to: 1) Characterise and contrast the phenotypes of different autism risk genetic variants, in terms of autism prevalence, severity, symptom domain profile, subdomain profile and IQ; 2) Investigate whether CNV carriers with autism differ in phenotype from individuals with autism of heterogeneous origins.

Methods

Participants

Genetics-first cohorts

We identified several clinical research sites and consortia which had independently established genetic-first cohorts, and had utilised the Autism Diagnostic Interview – Revised (ADI-R) [32] to assess autism, thus allowing data to be easily combined. Data on 566 clinically ascertained CNV carriers were available but 19 cases were removed due to insufficient genotypic information (n=18) and cohort overlap (n=1). This resulted in 547 CNV carriers (12.3 years (SD=4.2), 54% male) ; 82 with 16p11.2 deletion, 50 with 16p11.2 duplication, 370 with 22q11.2 deletion and 45 with 22q11.2 duplication was provided from the ECHO (ExperienCes of people witH cOpy number variants

https://www.cardiff.ac.uk/mrc-centre-neuropsychiatric-genetics-genomics/research/themes/developmental-psychiatry/echo-study-cnv-research) study, the IMAGINE-ID (Intellectual disability and Mental health: Assessing Genomic Impact on Neurodevelopment http://imagine-id.org/) study, the Hospital neurodevelopmental CNV cohort at the Belgrade University Children's Hospital Belgrade, the International 22q11.2DS Brain and Behaviour Consortium (http://22q11-ibbc.org/), the Center for Autism Research at Children's Hospital of Philadelphia, and the 16p11.2 European consortium (http://www.minds-genes.org/Site_EN/index.html) and Simons Variation in Individuals Project (VIP) Consortium (https://simonsvipconnect.org/) (Supplementary Table 1 has full details). Cohort demographics by continent are provided in Table 1. The characteristics of these studies have been described elsewhere[18,24,25,33-37].

Carrier status for CNVs at the 16p11.2 (critical region 29.6-30.2 Mb, had to span breakpoints 4 to 5, build hg19) or 22q11.2 (critical region 19.0-21.5 Mb, had to span at least low copy repeat regions A-B as pathogenicity of atypical variants outside the A-B region is uncertain, build hg19) loci was confirmed for all individuals through clinical chromosome microarrays, medical records and/or confirmation in a research laboratory (Full genotype information in Supplementary Table 2). Analysis included individuals ≥4 years old. The study was approved by the appropriate local ethics committees and institutional review boards. Each participant and his or her caregiver, where appropriate, provided informed written consent/assent to participate prior to recruitment.

Heterogeneous autism cohort

Data on 2053 individuals with autism, ≥4 years old, was accessed from the Autism Genome Project (AGP) [38]. These individuals were ascertained via autism diagnostic clinics. Of these 2053 individuals, 26 had CNVs at the 16p11.2 and 22q11.2 loci; 7 with 16p11.2 deletion, 4 with 16p11.2 duplication, 4 with 22q11.2 deletion and 11 with 22q11.2 duplication. Given the small sample sizes we did not compare these groups to the remainder of the AGP cohorts; also, previous work has reported on the

phenotype of CNV carriers in the AGP cohort[39]. These individuals were not included in the genomic condition groups given the different ascertainment strategies. The remaining 2027 individuals represent a group of individuals with autism for whom the underlying aetiology is heterogeneous (See Table 1 for demographics). Following previous authors[17], we refer to this cohort as " heterogeneous autism", rather than "idiopathic."

Autism assessment

All individuals were assessed using the Autism Diagnostic Interview - Revised (ADI-R)[32] by a research reliable assessor (further information on assessors and assessment sites in Supplementary Table 3). The ADI-R is a semi-structured interview conducted with the primary caregiver about a child's symptoms both currently and during early development. Total ADI-R score was used as an index of autism severity[37]. Autism domain scores for "Social Interaction", "Communication" and "Restricted, Repetitive, and Stereotyped Behaviours (RRBs)" were extracted, as well as autism subdomain scores (further details on ADI-R scores in Supplementary Materials). To meet autism criteria on the ADI-R an individual had to meet the clinical cut-offs on each domain (score of 10 for social, 8 (7 if nonverbal) for communication, and 3 for RRB) and there must also have been evidence of developmental abnormality before the age of 36 months.

Cognitive assessment

Full Scale IQ (FSIQ), Verbal IQ (VIQ) and Performance IQ (PIQ) scores were derived from age and developmentally appropriate standardized IQ measures as described elsewhere [18,24,33,35,40].

Statistical Analysis

Aim 1: Characterising and contrasting the phenotypes of different autism risk genetic variants

Autism prevalence within genetic variant groups

Autism prevalence was determined on the basis of the ADI-R diagnostic algorithm[41]. A logit mixed model was performed to determine whether genetic variant group (22q11.2 deletion, 22q11.2 duplication, 16p11.2 deletion, 16p11.2 duplication) was a predictor of autism diagnosis, whilst accounting for gender and age. Following previous international studies of the 16p11.2 duplication, we included site (European vs United States) as a covariate[18,42]. Post-hoc contrasts were conducted to establish autism prevalence differences between genetic variant groups with Tukey adjustment for multiple comparisons. The percentage of individuals who did not meet autism criteria but did meet the clinical cut-off in one or more domains was additionally calculated.

Autism profiles between genetic variant groups

To investigate possible differences in autism profiles between genetic variant groups, a series of analysis of covariance (ANCOVA) models were conducted with group as a predictor and the following phenotypic variables as outcome measures: ADI-R total as an index of *autism severity* (ADI-R total score)*, autism domain profile, autism subdomain profile* and *IQ profile*, whilst accounting for gender, age and site (see Supplementary Materials for full information). Tukey's method was used to conduct post-hoc contrasts between genetic variant groups, producing p-values adjusted for the number of contrasts. Eta-squared values were calculated to estimate the proportion of variance explained by genetic variant group (between group differences). We also calculated the variance that is explained by variable expressivity within the four genetic variant groups, i.e. variance not explained by genetic variant group, age, gender and site. Analyses of ADI-R total score, domain and subdomain scores were repeated including FSIQ as a covariate, to investigate whether differences in autism phenotype were driven by FSIQ.

Aim 2: Symptom profiles of individuals with autism within the genetic variant groups and individuals with "heterogeneous autism"

To compare autism in the genetic variant groups to *heterogeneous autism (*i.e., individuals from the AGP dataset who did not have 16p11.2 and 22q11.2 CNVs; n=2027), we conducted analyses leaving out individuals within the genetic variant groups who did not meet ADI-R criteria for autism, and compared the profiles to individuals with heterogenous autism. This resulted in 5 groups: *16p11.2 deletion + autism; 16p11.2 duplication + autism*; *22q11.2 deletion + autism; 22q11.2 duplication + autism*; and *heterogeneous autism* (Table 1).

MANCOVA analysis was conducted with group as a predictor and phenotypic scores as the outcomes, whilst accounting for gender, age and site. As in aim 1, analyses were run for *autism severity* (ADI-R total score)*, autism domain profile, autism subdomain profile* and *IQ profile*. Post hoc contrasts to investigate the difference between CNV + autism groups in relation to the autism group were conducted with Tukey adjustment for multiple comparisons.

To investigate whether male-to-female ratios differed between the five groups, we used a logit model with gender as a binary outcome, and group as a predictor, whilst taking account of fixed effects of age, and the random effect of site.

For aims 1 and 2 a Benjamini-Hochberg False Discovery Rate (B-H FDR) multiple testing correction of 0.05 was applied to p-values.

Results

Aim 1: Characterising and contrasting the phenotypes of different autism risk genetic variants

Autism prevalence within genetic variant groups

Within our cohort of CNV carriers ascertained clinically via medical genetics clinics and patient organisations; 43% of individuals with 16p11.2 deletion, 58% of individuals with 16p11.2 duplication, 23% of individuals with 22q11.2 deletion and 44% of individuals with 22q11.2 duplication met ADI-R criteria for autism (Table 1). Genetic variant group was a significant predictor of autism diagnosis

(p<0.001). Post-hoc contrasts revealed that autism prevalence in the 22q11.2 deletion carrier group (23%) was significantly lower compared to the 16p11.2 deletion (43%, p=0.004), 16p11.2 duplication (58%, p<0.001) groups; the remaining genetic variant group differences were not significant.

Within CNV carriers who did not meet formal autism diagnosis, we examined the proportion who met clinical cut-off criteria for one or more domains on the ADI-R. Amongst the 378/547 (69%) individuals who did not meet criteria for autism, 205/378 (54%) were found to meet the clinical cut-off for at least one domain, indicating a significant domain-based impairment; 38/47 (81%) of 16p11.2 deletion, 19/21 (90%) of 16p11.2 duplication, 135/285 (47%) of 22q11.2 deletion and 13/25 (52%) of 22q11.2 duplication carriers. Supplementary Table 4 and Supplementary Figure 1 show for each CNV the proportion of individuals who met the clinical cut-offs for each domain.

Autism profiles between genetic variant groups

Genetic variant group predicted autism severity (7% of the variance, p<0.001), autism domain profile (5% of the variance, p<0.001), autism subdomain profile (1% of the variance, p<0.001). In terms of individual domain scores, genetic variant group predicted 5% of the social domain total score, 3% of the communication domain score, and 15% of the RRB domain (Table 2, Figure 1). For subdomain scores the proportion of variance predicted by genetic variant group varied between 1% (social interaction) and 21% (motor mannerisms). In addition to motor mannerisms, the proportion of variance explained was also high for sensorimotor interests (19%). Genetic group variant predicted 12% of variance in FSIQ (p<0.001), 12% of variance in PIQ (p<0.001) and 4% of the variance in VIQ (p<0.001). Findings for autism severity, domain scores and subdomain scores remained significant after controlling for FSIQ, and the eta-squared values remained relatively unchanged (Supplementary Table 5). Age accounted for 0-3% of variance in phenotypic traits (see Supplementary Table 6).

After accounting for between group variability, age, gender and site, a large proportion of variability remained; 74% to 97% within group variability depending on trait (final column, Table 2). This is visualised in Figure 1 and Supplementary Figure 2, which shows that although group differences exist, there is much more variability within all groups across traits. For IQ we found greater variability for duplications than deletions for both 16p11.2 (Levene's test, $p=0.001$) and 22q11.2 (Levene's test, $p<0.001$) loci. For autism severity we found greater variability in outcome for duplications than deletions for the 22q11.2 locus (Levene's test, $p<0.001$) but not for the 16p11.2 locus (Levene's test, $p=0.071$).

Supplementary Table 7 details which post hoc Tukey contrasts between groups were significant (p-values adjusted for multiple contrasts). To briefly summarise phenotypic profiles; 16p11.2 deletion carriers had relatively moderate autism severity scores and moderate cognitive impairment (IQ= 81.3); 16p11.2 duplication carriers had relatively greater autism severity scores and greater cognitive impairment (IQ= 70.9); 22q11.2 deletion carriers had relatively lower autism severity scores but greater cognitive impairment (IQ=70.3); and 22q11.2 duplication carriers had relatively higher autism severity scores but less cognitive impairment (IQ=88.1).

Aim 2: Symptom profiles of individuals with autism within the genetic variant groups and individuals with "heterogeneous autism"

Supplementary Table 8 details mean scores for each phenotypic trait for each group (heterogeneous autism, 16p11.2 deletion + autism, 16p11.2 duplication + autism, 22q11.2 deletion + autism, 22q11.2 duplication + autism). With the exception of VIQ and "routines and rituals" domain, all phenotypic traits and subdomains were found to differ between the five groups (last column Supplementary Table 5). These findings remained significant after a B-H FDR 0.05 correction for multiple testing. Age accounted for 0-5% of variance in phenotypic traits (see Supplementary Table 9). Figure 2A visualises the profile of each "genetic variant + autism" group relative to each other and Figure 2B visualises the profile of each "genetic variant+ autism" group relative to the heterogeneous autism

group. Supplementary Table 8 details which aspects of the phenotypic profile showed significant contrasts between the heterogeneous autism group and the genetic variant groups.

To briefly summarise phenotypic profile differences relative to the heterogeneous autism group: the 16p11.2 deletion + autism group had relatively less impairment in autism score severity but had a similar level of cognitive impairment; the 16p11.2 duplication + autism group had greater PIQ deficits but did not differ on any of the other phenotypic measures; the 22q11.2 deletion + autism group had greater cognitive impairment but relatively less severity in autism scores; the 22q11.2 duplication + autism group did not significantly differ from the heterogeneous autism group on any phenotypic measure.

Sex

Male CNV carriers (all groups combined) were at increased risk of autism (OR = 2.3, $p<0.001$) compared to female CNV carriers. However, male to female ratios were lower within CNV carriers with autism (2.3:1) compared to the heterogeneous autism group (6.4:1) ($p<0.001$).

Discussion

This study is the result of a collaboration between several international genetics-first consortia and the Autism Genome Project. The availability of a large sample of individuals with one of four autism risk CNVs allowed us to use a genetics-first approach which meant we were not constrained by ascertaining patients on the basis of autism diagnosis, allowing examination of the impact of genotype on autism severity and domain profiles across the spectrum. The use of the widely accepted research diagnostic ADI-R interview across all sites represents a methodological strength, enabling us to directly compare the autism profiles of 22q11.2 and 16p11.2 CNVs. Our findings indicate that although autism risk genetic variants differ in several aspects of the autism phenotype, including autism severity, symptom domain profile and cognitive profile, only 1-21% of the variance is explained by genetic variant group, depending on autism measure. In contrast, variation within

each of the four genetic variant groups is much greater, explaining between 74%-97% of the variability, depending on autism measure. This highlights that even within individuals with the same autism risk genetic variant, the autism profile is difficult to predict on the basis of CNV alone and that phenotypic profiles overlap, providing evidence against a 'highly specific" model [43] whereby each genotype leads to a unique autism phenotype (see Supplementary Figure 3), instead our findings support a partially specific model whereby autism profiles are distinct but overlapping.

Severity of autism phenotype differed by genetic variant group. In terms of autism prevalence; fewer 22q11.2 deletion carriers met criteria for autism (23%) than 22q11.2 duplication (44%), 16p11.2 deletion (43%) and 16p11.2 duplication (58%) carriers. These figures represent autism prevalence within a clinically ascertained cohort of CNV carriers, and should not be taken as the prevalence for CNV carriers in the wider population. Among CNV carriers with autism we found that 22q11.2 deletion and 16p11.2 deletion carriers with autism had relatively less severe profiles compared to individuals with heterogeneous autism. On the other hand individuals with 16p11.2 duplication and 22q11.2 duplication with autism had a profile more consistent with individuals with heterogeneous autism. Our findings complement genome wide CNV studies which find the strength of association and penetrance for autism varies by genetic variant, in particular the association of 22q11.2 deletion is relatively weaker[8].

We found evidence that the four genetic variant groups were associated with differences in autism severity, the three autism domains as well as nine out of the 10 subdomains we studied, FSIQ, VIQ and PIQ. However, the proportion of variance explained by genetic variant group for each sub-domain varied between 1-21%. It was only the social interaction subdomain that did not differ, indicating that this trait is a universal aspect of autism across the 4 genetic variant groups. The sub-domains for which genetic variant group explained the greatest proportion of variance were motor aspects of the RRB domain, motor mannerisms (21%) and sensorimotor interests (19%), indicating that genetic variant group particularly distinguishes motor aspects of the autism phenotype.

Cognitive profile was also influenced by genetic variant group; 22q11.2 deletion and 16p11.2 duplication carriers had greater cognitive impairments in FSIQ, VIQ and PIQ relative to 22q11.2 duplication and 16p11.2 deletion carriers. There was evidence at both the 22q11.2 and 16p11.2 loci that cognitive outcomes are more variable for duplication carriers than deletion carriers. This has been previously been reported for 16p11.2 duplication carriers[18] and our findings indicate the same may be true for the 22q11.2 locus. Autism severity of a genetic variant did not covary with magnitude of cognitive deficit, 22q11.2 duplication carriers had the highest mean IQ (88.1) out of the CNV groups, yet had high symptom severity scores. 22q11.2 deletion carriers had the greatest cognitive impairment yet were at less risk of autism compared to the other genetic variants. Furthermore, when we controlled for IQ, differences in autism domain and subdomains scores between CNVs remained relatively unchanged. These findings suggest that the mechanisms underlying autism and cognitive impairment are at least partially distinct among carriers of pathogenic CNVs.

However, although specific group differences exist, it is clear that phenotypic profiles overlap (Supplementary Figure 2), and we find greater variability between individuals with the same CNV than between CNVs. Overall, our findings provide most support for a "partially specific model" whereby autism profiles are distinct but highly overlapping. Though the magnitude of these differences is closer to the "non-specific effect" end of the scale whereby all genotypes lead to similar autism phenotypes, than the "highly specific effect" end of the scale whereby genotypes lead to discrete autism subtypes (Figure 1). These findings highlight that it will be important for behavioural phenotyping research to move beyond a focus on average differences between variants, and to investigate the genetic (including additional rare variants and polygenic risk, which we were not able to analyse in this study) and environmental factors that contribute to variation in clinical phenotypes. There is already evidence that family background is important to consider in a genetic counselling context; parental IQ has been found to predict the IQ impairment in 16p11.2 and 22q11.2 deletion carriers[44-46].

There was a male preponderance for autism across all genetic variant groups, and gender significantly influenced domain and subdomain profiles. However, the male to female ratio in CNV carriers is approximately 2.3:1 which is considerably less pronounced than in the heterogeneous autism group (6.4:1). It may be that the genetic variants we studied have such a large effect on neurodevelopment that they partially override the protective effect of being female[9,11]. Age did influence phenotypic traits, however the proportion of variance age explained in analyses was low (≤5%).

Using a genetics-first approach, we identified a significant proportion of CNV carriers (54%) who did not meet autism criteria but did meet clinical cut-offs for diagnosis related impairments. Furthermore, the profile of CNV carriers with autism does to some extent present differently from heterogeneous autism (Figure 2B). This has the potential implication that the clinical needs of patients with genomic conditions may be overlooked because they fail to meet diagnostic criteria despite exhibiting a range of impairments across domains. Parents of children with CNVs at 16p11.2 or 22q11.2 who have taken part in our studies in the UK have anecdotally reported that their child's genetic diagnosis can be a barrier to receiving an autism diagnosis and support, with some service providers having stated that a child with a genetic diagnosis cannot also have a secondary diagnosis of autism despite DSM 5 specifying that autism can be diagnosed when "associated with a known medical or genetic condition or environmental factor"[1]. It is important that clinicians are aware of the risk of autism associated with certain genetic variants to improve the opportunities that these children receive of an early diagnosis and access to interventions.

Further clinical implications arise from our finding that there are not highly specific genotype-phenotype relationships between individual CNVs and autism, at least for 16p11.2 and 22q11.2 deletion and duplication variants. This indicates that, although CNVs are pre-symptomatically predictive of autism and therefore can inform early intervention, individual genotypes are not specific in predicting symptom subtypes. Rather our findings indicate an overlap in clinical

phenotypes between these CNVs, suggesting that neurodevelopmental service provision for different CNVs could be grouped together. Our genetics-first approach reveals great variability within CNV groups, highlighting that autism risk variants are not deterministic for autism. It is important that in genomic counselling that pathogenic CNVs are considered as one factor within a broader biopsychosocial context, rather than being the only causative factor for autism. Identification of genetic and environmental modifiers of phenotypes of autism risk CNVs has potential for informing clinical care and intervention.

Our study benefits from several features, including a large sample size by combining data from individuals with these rare genetic conditions from a number of international cohorts, and synchronisation of phenotyping measures across sites allowing for analysis extending beyond categorical diagnosis, allowing for autism domains and subdomains to be analysed. However, there are potential limitations. Firstly, ascertainment bias needs to be considered as our study focuses on individuals who received a clinical genetic diagnosis, and our findings therefore do not necessarily extend to individuals with these CNVs in the population who are affected below a clinical threshold and as a consequence not referred for genetic testing. As one of the main indications for genetic testing currently is often developmental delay[18,25], our findings may not be representative for individuals with these CNVs with a more typical developmental pattern. However, despite these ascertainment considerations, not all CNV carriers in this study met autism criteria or had cognitive impairment, thus allowing us to study the impact of genotype across a broad spectrum of abilities. Another source of possible ascertainment bias is that referral reason for genetic testing may differ by genetic variant. For instance it has been reported that 22q11.2 deletion carriers are more likely to be referred due to physical abnormalities, such as heart defects, in comparison to 22q11.2 duplication carriers who are more likely to be referred for developmental reasons[25]. However, this may actually reflect true phenotypic differences as a recent population based study which was able to identify individuals in the population undiagnosed with a 22q11.2 CNV, as well as individuals with a diagnosis through a clinic, reported higher frequency of congenital abnormalities in the deletion

carriers[47]. Before taking part in the study, individuals had a variety of diagnostic experiences, where some had a pre-existing autism diagnosis before the ADI-R assessment, whilst others had had no interaction with autism diagnostic services. This potentially introduces caregiver reporter bias, but this is partly mitigated by the semi-structured nature of the ADI-R. That is, although the ADI-R interview is based on caregiver report, the scoring of a particular trait is based on concrete descriptions coded by a trained interviewer. We were not able to conduct cross-site reliability of ADI-R administration as it was not pre-planned that ADI-R data would be combined across several international sites, however all assessors underwent ADI-R formal training and were research reliable. Finally, we were not able to control for ethnicity, and socio-economic and environmental factors, as these data were not available at all sites, and/or were not internationally comparable. Future studies would benefit from greater alignment of measurement of environmental factors across international sites.

Conclusion

The genetics-first approach we employed represents a novel method for investigating genotype-phenotype relationships unconstrained by categorical diagnostic criteria. We found that the phenotypic profiles of 16p11.2 and 22q11.2 CNVs differ in terms of severity, symptom profile and cognitive profile. However, although genetic variants have specific effects, within variant variability is much greater than between variant variability, thus indicating that the phenotypic consequences of genomic risk factors for autism fit a “partial specific model” rather than a “highly specific” model. It will be important that future studies of autism risk variants consider the genetic and environmental factors that contribute to clinical variability within autism risk variant carriers. An important message from our work is that individuals with genomic conditions are likely to present with clinically significant symptoms of autism but not meet diagnostic criteria. Clinical services need to adapt as individuals without a formal autism diagnosis are unlikely to access support and interventions.

Acknowledgements

We thank all the children and families who took part in this study, the UK National Health Service (NHS) medical genetic clinics for their support, and charities including Unique, Max Appeal and 22Crew for their support. We thank all members of the IMAGINE-ID consortium for their contributions. We thank the core laboratory team of the Division of Psychological Medicine and Clinical Neurosciences laboratory at Cardiff University (Cardiff, UK) for DNA sample management and genotyping. We are grateful to all of the families at the participating Simons VIP sites, as well as the Simons VIP Consortium. We appreciate obtaining access to the phenotypic data on SFARI Base. Approved researchers can obtain the Simons VIP (now known as Simons Searchlight) population dataset described in this study by applying at https://base.sfari.org. We also thank the International 22q11.2 Brain and Behaviour Consortium and the National Centre for Mental Health, a collaboration between Cardiff, Swansea and Bangor Universities, for their support.

Funding

This study was funded by the Baily Thomas Charitable Trust (MBMvdB, 2315/1), the Waterloo Foundation (MBMvdB, WF918-1234), Wellcome Trust ISSF grant (MBMvdB), the National Institute for Mental Health (MBMvdB, 5UO1MH101724; WRK, NIH / MH064824; CEB, RO1 MH085953; CEB and DMM-M U01MH101719; DMM-M U01MH087636), Medical Research Council studentship (SJRAC, 1499282), Wellcome Trust Fellowship (JLD, 505714), Wellcome Trust Strategic Award (MJO, 503147), Health & Care Research Wales (MJO, Welsh Government, 507556), Medical Research Council Centre grant (MJO, G0801418), Medical Research Council Programme grant (MJO, G0800509), the Simons Foundation (CEB, SFARI Explorer Award), Brain Canada foundation (SJ), Canadian Institute of Health Research (SJ, 400528 /159734, Canadian Institute of Health Research (SJ), Canada Research Chair in Genetics of Neurodevelopmental Disorders (SJ), Simons Foundation Autism Research Initiative (WKC), Swiss National Science Foundation (AM, grant PMPDP3_171331),

UK Medical Research Council and Medical Research Foundation grants (IMAGINE-ID study MBMvdB, JH & MJO; MR/L011166/1 and MR/N022572/1), and the collaboration with the University of Belgrade was funded by the European Cooperation in Science and Technology (MM, MBMvdB, SJRAC; COST action, CA16210). The authors thank the main funders of the Autism Genome Project: Autism Speaks (USA), the Health Research Board (Ireland; AUT/2006/1, AUT/2006/2, PD/2006/48), the Medical Research Council (UK), the Hilibrand Foundation (USA), Genome Canada, the Ontario Genomics Institute, and the Canadian Institutes of Health Research (CIHR).

Declaration of Interests

JH reports grants from Wyeth, Pfizer, AbbVie, A&Z Pharmaceutical, and Takeda Pharmaceuticals outside of the submitted work. MJO and MBMvdB report grants from Takeda Pharmaceuticals outside of the submitted work. All other authors declare no competing interests.

References

1 American Psychiatric Association. *Diagnostic and statistical manual of mental disorders, (DSM-5®)*. (American Psychiatric Pub, 2013).

2 Charman, T. *et al.* IQ in children with autism spectrum disorders: data from the Special Needs and Autism Project (SNAP). *Psychological medicine* **41**, 619-627 (2011).

3 Fountain, C., Winter, A. S. & Bearman, P. S. Six developmental trajectories characterize children with autism. *Pediatrics* **129**, e1112-e1120 (2012).

4 Kim, S. H., Macari, S., Koller, J. & Chawarska, K. Examining the phenotypic heterogeneity of early autism spectrum disorder: subtypes and short-term outcomes. *Journal of Child Psychology and Psychiatry* **57**, 93-102 (2016).

5 Eaves, L. C., Ho, H. H. & Eaves, D. M. Subtypes of autism by cluster analysis. *J. Autism Dev. Disord.* **24**, 3-22 (1994).

6 Lee, C. & Scherer, S. W. The clinical context of copy number variation in the human genome. *Expert reviews in molecular medicine* **12**, e8 (2010).

7 Pinto, D. *et al.* Functional impact of global rare copy number variation in autism spectrum disorders. *Nature* **466**, 368-372 (2010).

8 Malhotra, D. & Sebat, J. CNVs: Harbingers of a Rare Variant Revolution in Psychiatric Genetics. *Cell* **148**, 1223-1241, doi:http://dx.doi.org/10.1016/j.cell.2012.02.039 (2012).

9 Pinto, D. *et al.* Convergence of genes and cellular pathways dysregulated in autism spectrum disorders. *American journal of human genetics* **94**, 677-694, doi:10.1016/j.ajhg.2014.03.018 (2014).

10 Sanders, S. J. *et al.* Multiple recurrent de novo CNVs, including duplications of the 7q11. 23 Williams syndrome region, are strongly associated with autism. *Neuron* **70**, 863-885 (2011).

11 Levy, D. *et al.* Rare de novo and transmitted copy-number variation in autistic spectrum disorders. *Neuron* **70**, 886-897 (2011).

12 Sanders, S. J. *et al.* Insights into autism spectrum disorder genomic architecture and biology from 71 risk loci. *Neuron* **87**, 1215-1233 (2015).

13 D'Abate, L. *et al.* Predictive impact of rare genomic copy number variations in siblings of individuals with autism spectrum disorders. *Nature communications* **10**, 1-9 (2019).

14 Miller, D. T. *et al.* Consensus statement: chromosomal microarray is a first-tier clinical diagnostic test for individuals with developmental disabilities or congenital anomalies. *The American Journal of Human Genetics* **86**, 749-764 (2010).

15 Vorstman, J. A. *et al.* Autism genetics: opportunities and challenges for clinical translation. *Nature Reviews Genetics* **18**, 362 (2017).

16 Lord, C. & Veenstra-VanderWeele, J. Following the trail from genotype to phenotypes. *JAMA psychiatry* **73**, 7-8 (2016).

17 Bruining, H. *et al.* Dissecting the clinical heterogeneity of autism spectrum disorders through defined genotypes. *PLoS One* **5**, e10887 (2010).

18 D'Angelo, D. *et al.* Defining the effect of the 16p11. 2 duplication on cognition, behavior, and medical comorbidities. *JAMA psychiatry* **73**, 20-30 (2016).

19 Hanson, E. *et al.* The cognitive and behavioral phenotype of the 16p11. 2 deletion in a clinically ascertained population. *Biological psychiatry* **77**, 785-793 (2015).

20 Hippolyte, L. *et al.* The number of genomic copies at the 16p11. 2 locus modulates language, verbal memory, and inhibition. *Biological psychiatry* **80**, 129-139 (2016).
21 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* **2**, 909-916 (2015).
22 Vorstman, J. A. *et al.* The 22q11.2 deletion in children: high rate of autistic disorders and early onset of psychotic symptoms. *Journal of the American Academy of Child & Adolescent Psychiatry* **45**, 1104-1113 (2006).
23 Antshel, K. M. *et al.* Autistic spectrum disorders in velo-cardio facial syndrome (22q11.2 deletion). *J. Autism Dev. Disord.* **37**, doi:10.1007/s10803-006-0308-6 (2007).
24 Schneider, M. *et al.* Psychiatric Disorders From Childhood to Adulthood in 22q11.2 Deletion Syndrome: Results From the International Consortium on Brain and Behavior in 22q11.2 Deletion Syndrome. *American Journal of Psychiatry* **171.6**, 627-639 (2014).
25 Wenger, T. L. *et al.* 22q11. 2 duplication syndrome: elevated rate of autism spectrum disorder and need for medical screening. *Molecular Autism* **7**, 1 (2016).
26 Baird, G. *et al.* Prevalence of disorders of the autism spectrum in a population cohort of children in South Thames: the Special Needs and Autism Project (SNAP). *The lancet* **368**, 210-215 (2006).
27 Kogan, M. D. *et al.* Prevalence of parent-reported diagnosis of autism spectrum disorder among children in the US, 2007. *Pediatrics* **124**, 1395-1403 (2009).
28 Ronald, A. *et al.* Genetic heterogeneity between the three components of the autism spectrum: a twin study. *Journal of the American Academy of Child & Adolescent Psychiatry* **45**, 691-699 (2006).
29 Bruining, H. *et al.* Behavioral signatures related to genetic disorders in autism. *Molecular autism* **5**, 1 (2014).
30 Geschwind, D. H. & Levitt, P. Autism spectrum disorders: developmental disconnection syndromes. *Current opinion in neurobiology* **17**, 103-111 (2007).
31 Kates, W. R. *et al.* Comparing phenotypes in patients with idiopathic autism to patients with velocardiofacial syndrome (22q11 DS) with and without autism. *American Journal of Medical Genetics Part A* **143**, 2642-2650 (2007).
32 Rutter, M., Le Couteur, A. & Lord, C. *Autism Diagnostic Interview, Revised (ADI-R)*. (Western Psychological Services, 2003).
33 Niarchou, M. *et al.* Psychopathology and cognition in children with 22q11.2 deletion syndrome. *British Journal of Psychiatry* **204**, 46-54, doi:10.1192/bjp.bp.113.132324 (2014).
34 Chawner, S. *et al.* Genotype-phenotype relationships in children with copy number variants associated with high neuropsychiatric risk: Findings from the Intellectual Disability & Mental Health: Assessing the Genomic Impact on Neurodevelopment (IMAGINE-ID) study. *BioRxiv*, 535708 (2019).
35 Gur, R. *et al.* A neurogenetic model for the study of schizophrenia spectrum disorders: the International 22q11. 2 Deletion Syndrome Brain Behavior Consortium. *Molecular psychiatry* **22**, 1664 (2017).
36 Niarchou, M. *et al.* Psychiatric disorders in children with 16p11.2 deletion and duplication. *Translational Psychiatry* **9**, 8, doi:10.1038/s41398-018-0339-8 (2019).
37 A.M. Fiksinski, E. J. B. S. N. D. A. S. B. R. S. K. J. A. S. V. Autism Spectrum and psychosis risk in the 22q11.2 deletion syndrome. Findings from a prospective longitudinal study. *Schizophrenia Research* **0**, doi:10.1016/j.schres.2017.01.032 (2017).
38 Szatmari, P. *et al.* Mapping autism risk loci using genetic linkage and chromosomal rearrangements. *Nature genetics* **39**, 319-328 (2007).
39 Merikangas, A. *et al.* The phenotypic manifestations of rare genic CNVs in autism spectrum disorder. *Molecular psychiatry* **20**, 1366 (2015).

40 Szatmari, P. *et al.* Developmental trajectories of symptom severity and adaptive functioning in an inception cohort of preschool children with autism spectrum disorder. *JAMA psychiatry* **72**, 276-283 (2015).
41 Rutter, M., Le Couteur, A., Lord, C. & Faggioli, R. *ADI-R: Autism diagnostic interview--revised: Manual*. (OS, Organizzazioni speciali, 2005).
42 Niarchou, M. *et al.* Psychiatric disorders in children with 16p11. 2 deletion and duplication. *Translational psychiatry* **9**, 8 (2019).
43 Hodapp, R. M. Direct and indirect behavioral effects of different genetic disorders of mental retardation. *American Journal on Mental Retardation* **102**, 67-79 (1997).
44 Moreno-De-Luca, A. *et al.* The role of parental cognitive, behavioral, and motor profiles in clinical variability in individuals with chromosome 16p11. 2 deletions. *JAMA psychiatry* **72**, 119-126 (2015).
45 Klaassen, P. *et al.* Explaining the variable penetrance of CNVs: parental intellectual level modulates the expression of intellectual impairment caused by the 22q11.2 deletion. *Behavior and intelligence in children with the 22q11. 2-deletion syndrome*, 27 (2014).
46 Olszewski, A. K., Radoeva, P. D., Fremont, W., Kates, W. R. & Antshel, K. M. Is child intelligence associated with parent and sibling intelligence in individuals with developmental disorders? An investigation in youth with 22q11.2 deletion (velo-cardio-facial) syndrome. *Res Dev Disabil* **35**, 3582-3590, doi:10.1016/j.ridd.2014.08.034 (2014).
47 Olsen, L. *et al.* Prevalence of rearrangements in the 22q11. 2 region and population-based risk of neuropsychiatric and developmental disorders in a Danish population: a case-cohort study. *The Lancet Psychiatry* (2018).