Progress in the genetics of Autism Spectrum Disorder

Marc Woodbury-Smith\textsuperscript{1,2} and Stephen W. Scherer\textsuperscript{2,3,4}

Affiliations:
\textsuperscript{1}Institute of Neuroscience, Newcastle University, Newcastle upon Tyne, UK.
\textsuperscript{2}Program in Genetics and Genome Biology, The Centre for Applied Genomics, The Hospital for Sick Children, Toronto, ON, Canada.
\textsuperscript{3}Department of Molecular Genetics, University of Toronto, Toronto, ON, Canada
\textsuperscript{4}McLaughlin Centre, University of Toronto, ON, Canada.

Address for Correspondence:

Marc Woodbury-Smith,
Institute of Neuroscience,
Newcastle University,
c/o Sir James Spence Institute,
Queen Victoria Road,
Newcastle upon Tyne, UK NE1 4LP
Email: marc.woodbury-smith@newcastle.ac.uk
Tel: +44 (0)191-282-1392
Fax +44 (0)191-282-4725

Stephen W Scherer
The Centre for Applied Genomics,
Hospital for Sick Children,
13\textsuperscript{th} Floor,
Peter Gilgan Centre for Research and Learning,
686 Bay Street,
Toronto, ON, Canada M5S 1A1
Email: stephen.scherer@sickkids.ca
Tel: 1-416-813-7613
Fax: 1-416-813-8319
Abstract

A genetic basis for autism spectrum disorder (ASD) is now well established, and with the availability of high-throughput microarray and sequencing platforms major advances have been made in our understanding of genetic risk factors. Rare, often de novo, copy number and single nucleotide variants are both implicated, with many ASD-implicated genes showing pleiotropy and variable penetrance. Additionally, common variants are also known to play a role in ASD’s genetic etiology, but are yet to be identified. These new insights into the architecture of ASD’s genetic etiology offer opportunities for the identification of molecular targets for novel interventions, and provide new insight for families seeking genetic counselling.

Shortened title: Progress in genetics of ASD
What the paper adds

- A number of rare, often de novo, genetic variants, including deletions and duplications, are implicated in ASD, with some showing recurrence.
- Common genetic variants are also important and a number of associated loci are now being uncovered.
- Genetic testing for individuals diagnosed with ASD can offer the opportunity to identify relevant genetic etiology, thus facilitating counselling.

Introduction

Autism Spectrum Disorder (ASD) is a childhood onset lifelong disorder that impacts socio-communicative development and is also characterized by rigidity and ritualistic/repetitive patterns of behaviour. It occurs with a population prevalence of ~1%, with a male preponderance of 4:1. \(^1\) In approximately 50% of cases there is an association with intellectual disability (ID), and co-morbidity with neurodevelopmental and psychiatric disorders is frequent. \(^1\) In some cases, ASD forms part of a syndrome, usually in association with known single gene disorders. \(^2\) ASD can occur sporadically, but is often familial, with a sibling recurrence risk in the region of 10-20 times. \(^3\) Recognising these varied phenotypic manifestations, it is unsurprising that the underlying etiological architecture of ASD is also complex; it is widely accepted, however, that genetic factors play a crucial role, and that both common and rare forms of genetic variation confer susceptibility.

With the availability of affordable, high throughput microarray and whole genome sequencing (WGS) platforms (Figure 1), the spectrum of population based human genetic variation is now more fully understood. Individuals harbour in the region of 3 million genetic variants, 95% of which are shared with >5% of the population and are termed ‘common variants’, 4% shared with between 1-4% of the population, and 1% of which are rare or unique to any one individual or their immediate family. \(^4\) Some such variants are de novo, i.e. arise in the germline. This distinction between common and rare variation provides a useful
categorization for understanding the genetic architecture of many complex genetic disorders, including ASD. Specifically, the liability to ASD is now known to involve both rare variants of moderate effect size, and common variants, individually of small effect size but cumulatively conferring susceptibility above a theoretical level of liability. As discussed subsequently, much is known about the 'rare' end of the allelic spectrum of ASD’s etiology, with both de novo and rare inherited variants and complex structural rearrangements identified in both sporadic and inherited forms of the disorder. In contrast, although common variation is now known to be etiologically important, no replicated variants have yet been identified.

Figure 1 about here

**Rare mutations in ASD**

*Single gene disorders*: the identification of rare de novo and inherited genetic variants that are predicted damaging represents the most successful aspect of ASD gene discovery. Among some individuals, these risk variants form part of known genetic syndromes. Certain Mendelian genetic disorders are strongly associated with ASD, most notably Rett syndrome, among whom 40% are diagnosed with ASD, as well as Fragile X syndrome which is associated with ASD (25%) as well as other social phenotypes (e.g. social anxiety in >40%). Many other single gene disorders are also associated with ASD, including neurofibromatosis, tuberous sclerosis, and Williams-Beuren syndrome. All of these disorders are also associated with intellectual disability, and many with other neuropsychiatric disorders. Although individually these disorders are rare, cumulatively they account for approximately 10% of all cases of ASD. Routine screening for Rett syndrome and Fragile X syndrome among individuals with ASD has been established practice for several years, but clinicians should be alert to the possibility of other single gene disorders that may occur sporadically.

*Copy number variation*: with the availability of microarray platforms, offering higher resolution high-throughput genome scanning, duplication or deletion of segments of the genome have been described among individuals with ASD (Figure 1). These range in size from cytogenetically visible rearrangements to regions of copy number variation (typically involving more than 1kb) and smaller regions of insertion and deletion (indels, typically up to
Microarray studies largely converge on evidence for a greater burden for de novo mutations (both deletions and duplications) in ASD families compared with controls, with the proportion of de novo copy number variants (CNVs) three to five-fold higher in cases. Moreover, there is a higher mutational burden for large CNVs among ASD cases compared to non-ASD controls, as well as CNVs that overlap (i) known ID and/or ASD genes, (ii) pathogenic or otherwise clinically relevant genes, and (iii) genes that are brain expressed and more specifically those that are structurally and/or functionally related to the post-synaptic density (PSD) or chromatin remodelling/transcription regulation gene sets. At least 5% of individuals with ASD are carrying two or more penetrant mutations. Several rare recurrent, de novo exonic CNVs overlap genes that are promising candidates for ASD susceptibility, including NRXN1, PTCHD1, NLGN1 genes and SHANK1 and 3. Certain loci are recurrently found to be duplicated or deleted in ASD, including 16p11.2 (~0.8%), 15q11-13 (~0.5%) and 22q11 (~0.5%). Mutations in these genes and loci are also observed in the general population (Table 1, and see knowledge translation below). Other similar loci have been described, comprising small and large loci of CNV loss or gain. These loci are often pleiotropic and overlap DECIPHER genes listed in DECIPHER, a web-based platform for the sharing of plausibly pathogenic variants from well-phenotyped patients. They also as well as showing a similar association with other neuropsychiatric disorders. Although rare in the population, variants overlapping these regions are sometimes transmitted from parents who are without apparent phenotypic consequence. For example, exonic CNV deletions overlapping NRXN1 at 2p16.3 have been described in 0.11% of clinically ascertainment samples (0.11%), but are also seen in 0.02% of population controls. It is therefore important to consider the penetrance of such mutations when attempting to counsel families on risk. Additionally, smaller maternally inherited duplications enriched for CHD8 gene targets show a bias in transmission to ASD probands. It is possible that this class of inherited CNV of reduced penetrance (and SNV, see below) simply predisposes an individual to ASD, requiring additional factors (genetic or otherwise) to cause the clinical phenotype.

Table 1 about here

Single nucleotide variation (SNV): several large-scale sequencing projects have now been completed, including those that examine the entire coding region of the genome (whole
exome sequencing) and the entire genome (whole genome sequencing). Among simplex families, the rates of rare de novo and inherited SNVs that are predicted damaging is higher in ASD cases than in their unaffected siblings. In contrast, the rate of synonymous mutations does not differ. Protein truncating SNVs that impact conserved genes and show preferential maternal transmission are enriched in ASD probands. Similar to the maternally transmitted CNVs discussed above, these may be low penetrance mutations that predispose to ASD, requiring a ‘second hit’ to cause the clinical phenotype.

Strikingly, rates of exonic de novo mutations similar to these were also reported in a study of affected sib pairs; while in some cases an identical ASD-implicated mutation was shared between affected siblings, in others each sibling harboured unique ASD-implicated mutations. This evidence for genetic heterogeneity both between and within families, along with pleiotropy, speaks to the need for precision medicine, a personalized approach to genetics that is cognizant of the fact that assessment of an individual’s complete genomic risk is required.

One important finding from sequencing based studies is the relationship between paternal age and rates of rare, de novo variants. Specifically, the majority of point mutations in ASD originate from the father, with the correlation between paternal age and number of de novo events reflecting a 1.3-fold increase in the number of de novo events for every 10 years of paternal age. The evidence for a relationship between de novo mutations and maternal age is less consistent. A similar relationship is observed for de novo events in other neuropsychiatric disorders such as schizophrenia.

Common mutations in ASD

Among inherited forms of ASD, it is not unusual for first or second-degree relatives to share milder but related characteristics, termed the broader autism phenotype (BAP). In some cases, larger extended pedigrees show the segregation of such traits, and other neuropsychiatric diagnoses, more widely. A genetic underpinning for the tendency for ASD and related traits to run in families is supported by heritability estimates. The early twin studies were consistent with heritability in the region of 90%, higher if lesser, subclinical degrees of socio-communicative and other impairments were included. Although more recent studies have identified a larger role for shared environmental effects, this is likely
due to the overinclusion of concordant dizygotic twins along with different liability thresholds used in calculations. A heritability of 64% to 91% is supported if these potential confounds are taken into consideration. In intriguingly, autism traits are normally distributed in the general population, suggesting that clinical ASD may represent a threshold liability with underlying common polygenic variation conferring susceptibility. This is supported by the strong correlation between the polygenic contribution to ASD traits in the general population and among individuals with ASD. The fact that many families segregating ASD also seem to share risk for other neuropsychiatric disorders suggests that this genetic risk may be more correctly ascribed to fundamental brain-based traits, i.e. intermediate level phenotypes. There is a body of research examining these quantitative phenotypes; some have been shown to be heritable and associated with genome-wide linkage signals in families. In general terms, however, this has not led to gene discovery in ASD.

Until very recently, the search for common variants using genome-wide association (GWA) methods has also been largely unsuccessful. Several studies have been conducted, but in general terms, these identified signals that do not reach a genome-wide level of significance. Similarly, a recent GWA analysis of a discovery sample of 7,387 ASD cases failed to find individual variants that reached the genome-wide threshold; however, a meta-analysis with an independent sample (N=7,783 ASD cases) identified a genome-wide significant signal at 10q24.32 (rs1409313, OR=1.12, p=1.058e-08). Although this marker is intronic, it is in linkage disequilibrium with a number of other markers that span a gene rich region. The most recent GWAS study comprising 18,381 ASD cases and 27,969 controls identified five genome-wide significant signals, with a further seven loci shared with other traits also reaching a similar level of significance. Furthermore, evidence of heterogeneity of association across phenotypic subgroups was observed; for example, the higher functioning subgroup showed an excess of alleles that overlapped with educational attainment. Notwithstanding the negative findings, these and other GWA studies estimate the contribution of common variants to ASD’s heritability to be significant. For example, the recent meta-analysis described above estimated SNP-based heritability to be in the region of 32.6%. Similarly, the Swedish PAGES study (Population-based Autism Genetics and Environment Study) estimates heritability due to common variants to be in the region of
49.4%, with genetic liability spread evenly across chromosomes consistent with polygenic inheritance. The contribution of common variation seems to be the same regardless of IQ, and whether or not there is a background of de novo mutations. However, SNP heritability is higher among those cases without ID, supporting the observation that de novo and sporadic variants are more frequently observed among ASD cases with ID.

Knowledge translation

Interpreting results of genetic testing: One major challenge following genetic testing is the interpretation of results, and specifically, evidence of likely pathogenicity of identified variants. The American College of Medical Genetics and Genomics (ACMG) and the Association for Molecular Pathology (AMP) have issued joint guidelines for the classification of variants that uses terms ‘benign’ and ‘pathogenic’ and the strength of the evidence in favour of each. The framework utilizes information from the population frequency of variants, as well as computational in silico prediction and the results of functional studies that consider the impact of mutation on brain structure and function in vitro. In clinical terms, knowing the penetrance (i.e. the probability of a specified phenotype) of a particular mutation is important. The calculation of penetrance requires phenotyped population samples, which is often a problem considering the pleiotropy or variable expressivity associated with particular variants; subthreshold traits are likely to escape even the most detailed phenotyping. Penetrance can also be calculated by tracking inherited variants across family members, which is a particularly useful strategy for rare variants.

In general terms, although several genes are now identified as ASD genes, including NRXN1, SHANK 1 and 3 and PTCHD1, the penetrance of the underlying mutations is not well described. By way of example, deletions in NRXN1 are relatively common in clinically identified individuals with ASD (0.45%) and ID (0.12%), as well as in other neurodevelopmental phenotypes. In contrast, the frequency in population based surveys is much less (~0.02%). Consequently, an estimate for penetrance of NRXN1 mutations for all neurodevelopmental phenotypes is 33%. However, judging penetrance of conflated NRXN1 mutations does run the risk of overstating (or understating) penetrance for individual variants. This may be problematic for genetic counsellors who are required to make judgements about the risk of disease for an individual. Indeed, with regards NRXN1, most clinical cases
seem to cluster around exons 1-4 at the 5’ end of the gene, with deletions that impact the subsequent exons showing evidence of lower penetrance. 47

Individuals with an identical mutation may still demonstrate a large degree of phenotypic variability. By way of example, the 16p11.2 locus is associated with a range of neurodevelopmental disorders and other clinical phenotypes, including intellectual disability (ID). Haploinsufficiency at this locus shifts IQ downwards, but residual correlation with biparental IQ and ASD traits also suggest that background familial factors contribute to trait variability. 48 It is to be expected that a similar mechanism will explain phenotypic variation for other mutations. What is not entirely clear at present, however, is whether specific ASD symptoms are associated with particular genetic mutations. One recent analysis found that children with ASD who have de novo mutations had a relative strength in language but more pronounced motor deficits. 49 Similarly, among cases diagnosed with ASD, CNVs in ‘ASD/ID genes’ are specifically associated with communication sub-phenotypes. 50 Much remains to be learned about these phenotype-genotype relationships.

Providing genetic counselling for families: Genetic counsellors may be asked to advise on the recurrence risk of ASD among families who have one or more offspring already diagnosed with ASD. In families generally, the risk of having a child with ASD is related to the population prevalence, currently in the region of 1%. 1 Among those children deemed ‘high-risk’ by way of an older sibling being diagnosed with ASD, 18.7% are subsequently diagnosed with ASD. Moreover, with two older siblings with ASD, this figure goes up to 32.2%. 51

The situation is slightly different among those families where the first child already has been identified as harbouring a variant overlapping an ASD gene. For recurrent CNVs, for example, gene-specific and/or locus-specific penetrance calculations have been generated and may be useful in providing genetic counselling to families, particularly as they provide information about penetrance for ASD and other neurodevelopmental disorders. 52 Many variants, however, are categorized as ‘variants of unknown significance’. In such situations, the variant cannot be used for clinical decision making; instead, attempts to resolve variant classification are needed, and the family should be monitored. A more difficult and sensitive situation arises when genetic testing is being used for prenatal decision making. In such situations, the weight given to variant classification should be aligned with the strength of the
evidence, and in all likelihood other factors (including additional medical investigations and family circumstances) will play a bigger role in the decision-making process.

Assessment of an individual with ASD should include a thorough physical examination to document growth as well as body morphology. This may give light on the possible underlying genetic etiology. It is important to determine whether the individual has ‘complex’ or ‘essential’ ASD. The term ‘complex’ describes the existence of dysmorphology, abnormal growth, congenital anomalies and neurological complications. In contrast, such features are typically absent among those described as ‘essential’. Categorizing in this way facilitates the decision-making process in a clinical setting (Figure 2). For example, among individuals with head circumference >3SD from the mean the possibility of PTEN mutations should be considered, and developmental regression in a female patient should alert to the possibility of a MeCP2 mutation. For children with positive genetic test results, genetic counselling offers the opportunity to plan treatment and/or prophylactic measures for any medical conditions known to be associated with the particular gene. This may include close monitoring to prevent the onset of obesity in those with 16p11.2 mutations, mental health monitoring and intervention in those with 22q11 deletions and cardiovascular surveillance for those with 1q21.1 microdeletion syndrome.

Diagnostic yield of genetic testing: Children with developmental vulnerabilities are increasingly likely to have genetic testing. Among those who have ASD in association with dysmorphology and additional health problems, testing for Rett syndrome, Fragile X syndrome or one of the other single gene disorders is established practice. In recent years clinical microarray (CMA) has become feasible, and is increasingly used for children diagnosed with ASD. Indeed, a published consensus statement mandates the use of CMA as a first-tier diagnostic test in individuals with ID and/or ASD. The diagnostic yield of such testing is higher than G-banded karyotyping in clinical cohorts. For example, in one study of 848 children with ASD evaluated by CMA, 59 (7%) had results considered abnormal or possibly significant. Importantly, 83% of the abnormalities detected were below the size routinely detected by karyotyping. The yield of CMA has also been compared with exome sequencing. The yield from both were comparable (~8%), with both showing much higher
yield with complex ASD (21.9% vs 16.7% for CMA and exome sequencing respectively), defined according to the presence of dysmorphology as described in Miles et al. 57

**Development of molecular compounds:** The identification of ASD genes, and elaboration of their protein pathways, offers the opportunity for the development of new molecular compounds that target these protein networks. Recognising that large number of genes identified so far, and with estimates that up to 1000 genes may be involved, 58 it is hoped that these converge on a relatively small number of functional protein networks that may be amenable to treatment, although an ‘omnigenic’ model has also been proposed. 59 A number of clinical trials are currently underway, principally focussing on single gene disorder causes of ASD, such as Rett syndrome (IGF-1), Fragile X syndrome (mGluR5 antagonists) and tuberous sclerosis (mTOR inhibitors). 60 The confidence that these, and other, molecular compounds may impact on symptoms rests on the evidence for early plasticity of the brain. Indeed, it has been shown, for example, that phenotypes resulting from SHANK3 knock-out can be reversed in a mouse model in adult animals, 61 supporting potential therapeutic benefit during the period of plasticity and beyond. Similarly, inhibition of mTOR, a protein kinase, has reversed the brain changes associated with tuberous sclerosis in animal models, 62 and a reduction of mGluR5 signalling has resulted in symptomatic improvement in knockout mice. 61 The convergence of ASD-implicated genetic variants identified so far on specific cellular processes, most notably several proteins involved in synaptic structure and function, offers the opportunity to target drug development by focussing on one or more distal targets of the highly interconnected pathways involved. 64

**Conclusion**

In summary, major advances have been made in recent years in our understanding of the genetic architecture of ASD, with strong evidence that both rare and common variants contribute to risk, and with recent microarray and WGS platforms providing evidence to support a number of genes and loci as ASD-implicated. Despite this, however, much remains unknown: due to the infrequent nature of individual variants, their penetrance for ASD and related phenotypes is not well elaborated, and functional studies are only now beginning to contribute to the evidence in favour of, or against, pathogenicity. Several large-scale projects have now been initiated to help more fully realize the underlying complexity of ASD’s genetic architecture.
Acknowledgements

We thank The Centre for Applied Genomics (TCAG), which is funded by Genome Canada and the Ontario Genomics Institute, Canada Foundation for Innovation (CFI), and the Ontario Research Fund of the Government of Ontario. We thank the McLaughlin Centre, University of Toronto. S.W.S. holds the GlaxoSmithKline-CIHR Chair in Genome Sciences at the University of Toronto and The Hospital for Sick Children.

References


### Table 1: Example prevalences of rare CNVs in ASD and population samples

<table>
<thead>
<tr>
<th>Locus</th>
<th>Freq. in ASD</th>
<th>Freq. in population</th>
</tr>
</thead>
<tbody>
<tr>
<td>15q13.3del</td>
<td>0.002</td>
<td>0.002</td>
</tr>
<tr>
<td>15q13.3dup</td>
<td>0</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>16p11.2del</td>
<td>0.002</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>16p11.2dup</td>
<td>0.002</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>NRXN1 del</td>
<td>0.002</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>SHANK3 del</td>
<td>0.002</td>
<td>NK</td>
</tr>
</tbody>
</table>

* Frequency based on MSSNG for ASD probands (N=1,739) and affected siblings (N=878).
** Frequency based on UK Biobank samples (N=152,000).

del=deletion; dup=duplication; nk=unknown
Legend

**Figure 1:** Genotyping platforms with details of their resolution, the types of variants detected and example ASD loci that have been identified.

**Figure 2:** Flow diagram of approach to genetics investigation and evaluation for individuals with ASD