Title
Genetics of autism spectrum disorder.

Permalink
https://escholarship.org/uc/item/9zn03530

Journal
Handbook of clinical neurology, 147

ISSN
0072-9752

Authors
Ramaswami, Gokul
Geschwind, Daniel H

Publication Date
2018

DOI
10.1016/b978-0-444-63233-3.00021-x

Peer reviewed
Chapter 21
Genetics of autism spectrum disorder

GOKUL RAMASWAMI1 AND DANIEL H. GESCHWIND1,2,3*

1Program in Neurogenetics, Department of Neurology, David Geffen School of Medicine, University of California, Los Angeles, CA, United States
2Center for Autism Research and Treatment, Semel Institute, David Geffen School of Medicine, University of California, Los Angeles, CA, United States
3Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, CA, United States

Abstract
Autism spectrum disorder (ASD) is a prevalent neurodevelopmental disorder characterized by impaired social interaction and stereotyped behaviors. ASD has a strong and complex genetic component, with multiple familial inheritance patterns and an estimate of up to 1000 genes potentially implicated. Over the past decade, genomic technologies have enabled rapid progress in the identification of risk genes for ASD. In this chapter, we review the delineation of ASD disease genes starting from traditional genetic studies such as linkage and association, and then focusing on more recent studies utilizing genomic technologies, such as high-throughput genotyping and exome sequencing.

INTRODUCTION

Autism spectrum disorder (ASD) is a heterogeneous neurodevelopmental disorder mainly characterized by impaired social interaction and restricted and stereotyped behaviors (American Psychiatric Association, 2013). ASD has an incidence of about 1 in 68 individuals and a 3:4 ratio of affected males to females (Elsabbagh et al., 2012). Associated comorbid phenotypes include other neurodevelopmental conditions such as intellectual disability, neurologic disorders such as epilepsy, soft neurologic motor signs, and nonneurologic medical conditions such as gastrointestinal and cardiac problems (Geschwind, 2009; Aldinger et al., 2015). There exists a large degree of clinical heterogeneity and developmental trajectories underlying ASD (Szatmari et al., 2015), even between monozygotic twins (Kolevzon et al., 2004), suggesting roles for both genetic and environmental factors.

ASD has a large heritable component. Infants born to families with an affected older sibling have a 20% chance of developing ASD (Ozonoff et al., 2011). Concordance rates among monozygotic twins, dizygotic twins, and siblings are 30–99%, 0–65%, and 3–30%, respectively, with an estimated overall heritability of 0.7–0.8 (Bailey et al., 1995; Rosenberg et al., 2009; Hallmayer et al., 2011; Colvert et al., 2015). However, recent population-based studies in Scandinavia show a lower estimated heritability of 0.5–0.6 (Sandin et al., 2014), which is not surprising given the wide confidence intervals in the published heritability estimates based on twins. Additionally, over the last decade the discovery of a significant role for genetic, but nonheritable, de novo mutations in ASD risk has emerged. The contribution from de novo mutations is estimated to be between 15% and 25% (Ronemus et al., 2014), indicating that in total genetic factors account for a large fraction of ASD risk in the population. Environmental factors are thought to play a minor role; however, they have not been well characterized (Newschaffer et al., 2007), with only a few examples such as prenatal exposure to valproate (Christensen et al., 2013) linked to increased risk of ASD.

*Correspondence to: D.H. Geschwind, 2506A Gonda Bldg, 695 Charles E. Young Dr. South, Los Angeles CA 90095-1761, United States. Tel: +1-310-794-6570, E-mail: dhg@mednet.ucla.edu
Although the overall genetic liability for ASD is high, its genetic architecture is extremely diverse with contributions from alleles of varying frequencies (i.e., common, rare, very rare), inheritance patterns (i.e., dominant, recessive, X-linked, de novo), and variant type (i.e., large chromosomal rearrangements, copy number variants (CNVs), small insertions/deletions (indels), and single-nucleotide variants (SNVs)) (De Rubeis and Buxbaum, 2015; de la Torre-Ubieta et al., 2016). Early studies using classic genetic mapping techniques such as linkage, candidate gene screening, and common variant association provided modest results, mainly because they were underpowered to detect causal variants in light of the high genetic heterogeneity in ASD. Over the past decade, advances in DNA sequencing technologies along with the collection of large patient cohorts and widespread data sharing have enabled enormous progress in the identification of genes implicated in ASD. Over 100 rare Mendelian syndromes have been associated with ASD (Abrahams and Geschwind, 2008; Betancur, 2011), indicating that, at least in part, ASD is a grouping of rare disorders. In contrast to the strong male bias in idiopathic ASD, syndromic ASD cases have a 1:1 male-to-female ratio (Yoo, 2015). Classic neurodevelopmental syndromes such as fragile X syndrome (caused by mutations in FMR1), Rett syndrome (caused by mutations in MECP2), tuberous sclerosis (caused by mutations in TSC1 and TSC2), Timothy syndrome (caused by mutations in CACNA1C), and dup15q syndrome all display partial comorbidity with ASD. Mutation analysis in multiplex families has been a fruitful strategy to identify genes underlying syndromic ASD. Rare X-linked mutations in two members of the neuregulin family NLGN3 and NLGN4 were identified in males diagnosed with ASD and mental retardation from multiple families (Jamain et al., 2003). Using linkage analysis or homozygosity mapping in consanguineous families, rare recessive mutations were identified in CNTNAP2 from Amish families as well as in SLC9A9 and BCKDK from Middle Eastern families in individuals with ASD and epilepsy (Strauss et al., 2006; Morrow et al., 2008; Novarino et al., 2012). Rare likely damaging inherited variants have also been identified in three members of

**COMMON VARIANTS**

The contribution of common variants to ASD has been mainly assessed using genome-wide association studies (GWAS). GWAS utilize high-throughput genotyping of common variants, defined as genetic polymorphisms present in at least 5% of the population. Common variants are estimated to play an important role, accounting for 40–60% of the overall liability of ASD (Klei et al., 2012; Gaugler et al., 2014). However, the effect size of each individual variant is quite small (Anney et al., 2012), necessitating large sample sizes to achieve sufficient detection power to identify specific risk alleles. To date, all of the GWAS studies in ASD have been underpowered, with sample sizes on the order of a few thousand cases (Ma et al., 2009; Wang et al., 2009; Weiss et al., 2009; Anney et al., 2010, 2012; Chaste et al., 2015). In contrast, GWAS of other prevalent diseases such as schizophrenia have sample sizes on the order of 30,000 cases (Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014). For ASD, only two loci have reached genome-wide significance: intergenic variants between CDH9 and CDH10 on chromosome 5p14.1 (Wang et al., 2009) and variants within the MACROD2 gene on chromosome 20p12.1 (Anney et al., 2010). Neither has been replicated in an independent study. For robust detection of common ASD risk alleles, future GWAS studies will require much larger sample sizes to be sufficiently powered.
a family of scaffolding proteins, SHANK1, SHANK2, and SHANK3, as well as two members of the neurexin family, NRXN1 and NRXN3 (Moessner et al., 2007; Kim et al., 2008; Berkel et al., 2010, 2012; Sato et al., 2012; Vaags et al., 2012), indicating a potential role of these genes in ASD.

The emergence of exome sequencing (DNA sequencing of all protein-coding regions of the genome) has enabled high-throughput assessment of protein-disrupting mutations in thousands of individuals. New syndromes associated with ASD continue to be identified via exome sequencing, such as a combination of facial dysmorphism and ASD caused by mutations in CHD8 (Kumar et al., 2008; Sanders et al., 2015), indicating a potential role of these genes in ASD.

Exome sequencing studies of large ASD cohorts have helped elucidate the contribution of inherited alleles to the risk for ASD by identifying an increased rate of rare inherited truncating mutations in genes that are infrequently mutated in the population (Lim et al., 2013; Knunn et al., 2015). Exome sequencing has also been applied to consanguineous families to identify inherited recessive variants in the genes AMT, PEX7, SYNE1, VPS13B, PAH, and POMGNT1 (Yu et al., 2013). Future exome and whole-genome sequencing (WGS) of multiplex families, such as those in the Autism Genetic Resource Exchange, will substantially expand knowledge of rare inherited alleles in ASD (Lajonchere and Consortium, 2010).

**DE NOVO VARIATION**

De novo variants arising in the parental germline are the underlying cause for many cases of ASD. The first de novo events observed in ASD were large CNVs. Using cytogenetic techniques, large deletions or duplications, such as duplication of 15q (Bundey et al., 1994), deletion of 22q11.2 (Fine et al., 2005), deletion of Xp22.3 (Thomas et al., 1999), and duplication or deletion of 16p11.2 (Weiss et al., 2008), were identified as genetic risk factors for ASD. During the past decade, oligonucleotide arrays for high-throughput genotyping have enabled detection of submicroscopic CNVs. Using array comparative genomic hybridization, Sebat and colleagues (2007) demonstrated that de novo submicroscopic CNVs were strongly associated with ASD. It quickly became apparent that ASD patients harbor significantly more de novo CNVs than control subjects (Levy et al., 2011) and that multiple recurrent de novo CNVs, such as duplication of 7q11.23 and microdeletion of 16p11.2, are prevalent in ASD patients (Kumar et al., 2008; Sanders et al., 2011). De novo CNVs exhibit vast heterogeneity in size – some are small enough to affect a single gene or large enough to encompass many genes – and clinical presentation; patients with multiple de novo CNVs usually have more severe phenotypes (Marshall et al., 2008; Pinto et al., 2010). More recent studies with larger cohorts, such as the Simons Simplex Collection, a collection of ASD families with a single affected child and usually one or more unaffected siblings, have found that genes within disease-associated de novo CNVs converge upon particular cellular functions such as neuronal signaling, synaptic function, and chromatin remodeling (Pinto et al., 2014; Sanders et al., 2015).

Exome sequencing studies in simplex cohorts, such as the Simons Simplex Collection, have identified numerous de novo variants implicated in ASD. In 2012, four groups published studies using exome sequencing on trios to identify many novel ASD susceptibility genes harboring de novo loss of function or likely gene-disrupting SNVs (Iossifov et al., 2012; Neale et al., 2012; O’Roak et al., 2012b; Sanders et al., 2012). Some of the recurrent hit genes were DYRK1A, POGZ, CHD8, NTNG1, GRIN2B, KATNAL2, and SCN2A (Chen et al., 2015). All of these studies found that de novo SNVs have a predominantly paternal origin and the rate of de novo mutation frequency increases with paternal age, concordant with other studies on de novo mutation frequencies (Kong et al., 2012). Subsequent exome sequencing studies have identified a number of genes implicated in ASD (Table 21.1) (De Rubeis et al., 2014; Iossifov et al., 2014). Although the genes identified so far are functionally heterogeneous, many of them are significantly involved in synaptic formation, transcriptional regulation, and chromatin remodeling (De Rubeis et al., 2014; O’Roak et al., 2012a, 2014). It is estimated that there are 400–1000 ASD risk genes (Geschwind and State, 2015).

WGS is starting to be utilized to comprehensively identify all de novo variants in ASD families. In contrast to exome sequencing, WGS can identify mutations in non-coding regulatory elements as well as coding mutations. Although it is currently challenging to interpret the functional impact of noncoding mutations, information such as evolutionary conservation, functional genomics, chromatin state, sequence motifs, and molecular quantitative trait loci is currently being integrated to facilitate annotation of noncoding genomic regions (Ward and Kellis, 2012). Recently, the first study using WGS to identify potentially disease-relevant noncoding variants in simplex ASD families found a significant enrichment of de novo and private disruptive mutations in DNAse I hypersensitivity sites located within 50 kb of genes previously implicated in ASD by exome studies (Turner et al., 2016).
GENETIC CONTRIBUTIONS TO ASD IN THE POPULATION

The contributing genetic alleles have not been delineated in a majority of patients with ASD. Although hundreds of ASD risk genes have been identified, the contribution of each particular gene within the ASD population is very small, with none found in more than 2% of patients. The largest class of genetic risk for ASD, around 50% of the total, is estimated to derive from common variants of an additive effect, almost all of which have yet to identified (Gaugler et al., 2014). A further 5–10% of genetic risk is estimated from rare inherited and de novo variants, leaving 40% of the genetic risk currently unaccounted for (Fig. 21.1).

CLINICAL GENETIC TESTING

Currently, yields for detection of large structural abnormalities by clinical microarray testing in all patients is 5–10% and fragile X testing in males is 2%, making these tests part of the standard of practice (Schaefer et al., 2013; Jeste and Geschwind, 2014; Tamminies et al., 2015; de la Torre-Ubieta et al., 2016). Even higher yields for clinical diagnosis using exome sequencing are expected based on the published research studies (Iossifov et al., 2014) as well as reports of clinical cohorts (Lee et al., 2014). In neurodevelopmental disorders, the yield of exome sequencing is about 30% (Lee et al., 2014), making it among the most powerful diagnostic tests available. Moreover, in particular instances, genetic diagnosis by exome sequencing has directly informed therapeutic strategies. In branched-chain keto acid dehydrogenase kinase deficiency (OMIM 614923), a syndromic form of ASD, identification of inherited recessive truncating mutations in BCKDK, a negative regulator of branched-chain amino acid (BCAA) degradation, implicated that a deficiency of BCAAs in these patients contributed to their pathogenesis (Iossifov et al., 2014). Common GO terms were identified as those shared by three or more genes. The resulting list was shortened using REVIGO (Supek et al., 2011) to obtain a representative subset of GO terms using an algorithm that clusters by semantic similarities, avoiding redundancy.
diagnosis of ASD. Dietary supplementation of BCKDK knockout mice with increased BCAAs led to an amelioration of neurologic phenotypes, suggesting a promising therapy for ASD patients with BCKDK mutations (Novarino et al., 2012). Thus, genetic testing first by microarray, followed by exome sequencing or WGS, should be integrated into the standard of practice for ASD (Fig. 21.2).

CONCLUSIONS
Empowered by genomic technologies and large data-sharing efforts, human genetic studies are making rapid progress in identifying risk genes for ASD. Despite the large clinical and genetic heterogeneity underlying ASD, genotype–phenotype correlations are starting to be identified by clustering patients based on their genetic background, a so-called genotype-first approach (Stessman et al., 2014). The integration of genetics with clinical phenotypes and other functional genomic techniques such as transcriptomics and epigenomics will lead to a better understanding of the molecular mechanisms involved in ASD and ultimately inform clinical care.

ACKNOWLEDGMENTS
We thank Laura Perez-Cano and Elizabeth Ruzzo for critical reading of the manuscript. Funding was provided by grants from National Institutes of Health (5R01 MH100027; 5P50 HD055784-09; 1R01 MH094714-05; 5R01 NS073871-04; 5R01 MH100900-02) and California Institute for Regenerative Medicine (GC1R-0667-A).
### References


GENETICS OF AUTISM SPECTRUM DISORDER


