Autism spectrum conditions (henceforth ‘autism’) refer to a group of neurodevelopmental conditions involving difficulties in social interaction and communication and unusually repetitive and restricted behaviours and interest. Twin and family studies have established a significant heritability for autism. Autism is polygenic with variations across the allele frequency spectrum contributing to risk. Early linkage and candidate gene association studies were statistically underpowered to identify significant loci. Current genome-wide association studies have identified significant positive genetic correlation between autism and various measures of cognition. The use of genetic microarrays and next-generation DNA sequencing has identified tens of genes and copy number variants associated with autism. In addition, RNA microarray and sequencing studies of postmortem brain samples have identified transcriptionally altered genes and pathways in autism. Multiple lines of evidence converge on altered glial, synaptic and chromatin pathways as contributing to autism risk.

Introduction

Autism spectrum conditions (henceforth autism) refer to a heterogeneous group of neurodevelopmental disorders characterised by difficulties in social interaction and communication alongside unusually repetitive behaviour and unusually narrow interests (Lai et al., 2013). It has an estimated prevalence of 1 in 100, although recent reports from the Centers for Disease Control and Prevention (CDC) suggest that it may be higher (1 in 68). The median worldwide prevalence is estimated between 0.62 and 0.70 (Lai et al., 2013). There is a marked sex difference: between 3 and 4 times as many males are diagnosed with autism compared to females, which may be due to innate biological factors (e.g. a female protective effect) or due to other social or cultural factors (e.g. females with autism may be better at ‘masking’ their phenotypes and hence may not be diagnosed with autism, and females are often diagnosed later than males) (Lai et al., 2013). In addition to the difficulties in the core diagnostic criteria, individuals with autism often have other comorbid phenotypes. For example, many individuals have atypical language development, sensory hypo or hypersensitivity and difficulties in motor coordination, including dyspraxia. Approximately 38% of all individuals with autism have intellectual disability (ID), as estimated by the CDC. Similarly, there is a reasonably high comorbidity with ADHD, depression and suicidal ideation (Lai et al., 2013).

Aside from clinical comorbidities, on average, individuals with autism tend to perform better on measures of ‘systemising’ drive (Baron-Cohen et al., 2003), that is, the drive to analyse and build systems, based on identifying the laws that govern the particular system, in order to predict how that system will work. Systems may be abstract, mechanical, natural, collectible and motoric. They also perform better on tests of attention to detail (Jolliffe and Baron-Cohen, 1997), a prerequisite for systemising. Individuals who are in the science-technology-engineering-maths (STEM) fields, or relatives of these individuals, are more likely to be diagnosed with autism or have higher levels of autistic traits (Ruzich et al., 2015; Wheelwright and Baron-Cohen, 2001; Baron-Cohen et al., 1997). Autistic individuals also, on average, tend to have difficulties in eye contact (Jones and Klin, 2013), attention to social stimuli (Dawson et al., 1998) and interpreting emotions (Baron-Cohen et al., 1997), which contribute to persistent difficulties in social interaction and communication. This may be because the social domain is less amenable to systemising, as it does not reduce to a set of rules.

There is considerable evidence that autism is partly genetic. Early twin studies established a prominent role for genetic variants by investigating concordance rate of both the clinical phenotype and associated behavioural and cognitive phenotypes in twins (reviewed in Bourgeron, 2016a; Ronald and Hoekstra, 2011). Recently, with increase in diagnosis rates and the availability of electronic records, it has been possible to establish familial recurrence rates in large population-based cohorts, providing converging evidence for familial recurrence rates (Sandin et al., 2014; Sandin et al., 2015). Buoyed by the heritability estimates, early molecular genetic studies focused largely on linkage...
and candidate gene sequencing and association studies in autism. Candidate gene association studies were conducted in relatively small sample sizes, with systematic inflation of effect sizes (Warriner et al., 2015a). Most of the results from these studies have not replicated in studies with larger samples.

However, there were some early successes in candidate gene sequencing studies, particularly in identifying genes for syndromic conditions with comorbid autism such as FMR1, TCS2 and NFI1 (Bourgeron, 2016a). The considerable genetic and phenotypic heterogeneity in autism has made it challenging to identify genes associated with the condition, necessitating investigations in large population cohorts given the small effect sizes and/or rarity of the genetic variants. Recent studies have leveraged large sample sizes available in population databases for autism such as the Simon’s Simplex Collection and the Autism Genetic Research Exchange. In this article, we review the different methods used to investigate the genetic architecture of autism and related traits, and recent progress in autism genetics research.

Establishing heritability: twin studies and familial recurrence

Since the first description in 1943 by Kanner (1943), autism was known to be a condition that manifested in early childhood, leading to the hypothesis that the condition is at least partly genetic. Establishing heritability is important, as it provides evidence for a causal role of genes in autism risk, which can then inform molecular genetic studies. Early family studies that were conducted in the 1960s and 1970s did not find evidence for familiality, which was because autism was perceived as an extremely rare condition (in the 1960s and 1970s, prevalence estimates for autism were 2–4 in 10,000). However, deeper examination of family factors provided two clues that autism is likely to be, at least partly, heritable. First, the recurrence rate in siblings was considerably higher than the risk in the general population, and second, there was a family history of delayed speech in about a quarter of the families of autistic individuals surveyed. Despite this, early twin studies in the 1960s and 1970s were inconclusive due to methodological issues and the lack of statistical power.

The first twin study to report evidence for familiality in autism, by Folstein and Rutter in 1977, investigated the concordance of autism in a small sample of 11 monozygotic twins (MZ) and 10 dizygotic twins (DZ) (Folstein and Rutter, 1977). The concordance for autism was 36% in the MZ twins and 0% in DZ twins. Expanding the criteria to include associated cognitive and social impairment or the broad autism phenotype (BAP) showed that 82% of the MZ twins were concordant, whereas only 10% of the DZ twins were concordant. Since then, several studies have investigated the heritability of autism in twin samples in different populations. Heritability estimates have largely been comparable and high across twin studies, regardless of the ascertainment criteria (Ronald and Hoekstra, 2011).

A recent meta-analysis of seven twin studies identified a high twin heritability of 64–91% (Tick et al., 2016). In parallel, twin studies of ‘autistic traits’ have identified modest to highheritabilities between 60% and 90%, although this varies depending on the type of measure used and the age of the participants (Ronald and Hoekstra, 2011). A few studies have also conducted multivariate coheritability analyses of autism and related traits, suggesting a significant shared genetic influence. The genetic correlation between autism/autistic traits and ADHD/ADHD traits in particular is high reflecting the phenotypic comorbidity (Ronald and Hoekstra, 2011). See also: Twin Studies

Family recurrence rates have also provided evidence for heritability for autism. Multiple studies from Scandinavian countries have identified similar risk ratio for siblings of individuals with autism (~10%) (Sandin et al., 2014; Grønborg et al., 2013; Jokiranta-Olkoniemi et al., 2016). Family recurrence rates also offer other clues into the underlying genetic architecture of autism. One interesting observation is that siblings of female probands have higher risk for autism than siblings of male probands – an observation that is called the Carter effect (the effect was originally described in pyloric stenosis (Carter and Evans, 1996), but subsequently used in other conditions). The Carter effect suggests that females have a protective effect, suggesting that a greater mutation burden is required for a clinical diagnosis of autism (this has been confirmed using gene sequencing studies, see the following discussion). If the genetic risk is partly familial, that is not de novo, this higher genetic risk can be inherited by siblings, thus increasing the risk for autism. Investigations of twin and multiplex family samples have identified a higher relative risk for autism in siblings of female probands (Werling and Geschwind, 2015; Robinson et al., 2013). In contrast, large population studies have failed to find support for the Carter effect (Sandin et al., 2014).

Another interesting finding is the identification of autistic traits or the BAP in family members of probands (Wheelwright et al., 2010). This is also supported by the identification of higher relative risk for other psychiatric conditions in family members of probands compared to the general population (Frazier et al., 2015; Constantino et al., 2010; Jokiranta-Olkoniemi et al., 2016).

A third interesting finding is that of parental age. A few studies have demonstrated that increased paternal age increases the risk for autism (Sandin et al., 2015; Frans et al., 2013; McGrath et al., 2014). Indeed, a considerable proportion of de novo mutations are paternal in origin (Kong et al., 2012; Gratten et al., 2016). Earlier studies have noted that increased paternal age also increases the risk for de novo mutations in the sperm as the number of cell replications increases with age (Kong et al., 2012; Gratten et al., 2014).

Finally, another hypothesis which has received considerable support is the idea that higher autistic traits or psychiatric liability in individuals is likely to delay fatherhood (McGrath et al., 2014; Gratten et al., 2014). However, a recent study has shown in addition to increased paternal age, the difference in ages of the parents also contributes to risk for autism, with the risk increasing as the difference in ages increases (Sandin et al., 2015). The mechanism underlying the increased risk for increased differences in parental age is unclear.

Early studies: linkage

With evidence of considerable heritability for autism from twin studies, early genetic studies focused on linkage of multiplex autism families to identify loci associated with the condition. Linkage studies investigate the inheritance of regions of
chromosomes in family pedigrees. The first autism linkage study was reported by IMGSAC in 99 families (International Molecular Genetic Study of Autism Consortium, 1998). Early linkage studies were nonparametric and in a relatively small sample size of 100–200 families. These studies used relatively sparse linkage genotyping and yielded limited success. Although several loci have been associated with autism or related traits, only a few of these loci have been consistently associated with the condition (Geschwind and State, 2015a). Some of these loci overlap with observed copy number variation in autism, suggesting that these may indeed reach statistical significance as sample sizes increase. For example, Alarcón and colleagues identified linkage at 7q35 when investigating language-related quantitative traits in autism (Alarcón et al., 2008). Copy number variation in 7q35 has also been implicated in autism (see: https://gene.sfari.org/autdb/CNVSecDis.do?l=7q35), although this is not significant at a genome-wide level (Sanders et al., 2015). Replication and functional analysis in foetal postmortem brain tissues identified CNTNAP2 as a likely candidate gene associated with language-related phenotypes in autism (Alarcón et al., 2008). However, subsequent sequencing studies have failed to identify a significant association between rare genetic variants in CNTNAP2 and autism (Murdoch et al., 2015; Sanders et al., 2015), although Cntanp2 mice display language and social deficits (Peñagarikano et al., 2011). See also: Linkage Analysis

Early studies: syndromic forms of autism and candidate gene associations

Many genes currently associated with autism were first identified through specific syndromic forms of autism. For example, four early genes identified with autism – FMR1, TSC1 and TSC2 and MECP2 – are all associated with syndromic forms of autism (fragile X syndrome, tuberous sclerosis and Rett syndrome). Early studies also investigated large chromosomal abnormalities by karyotyping. Because of the low resolution of these studies, it was impossible to identify specific genes associated with the condition. Today, the prevalence of chromosomal abnormalities is estimated to be less than 2% in autism (Bourgeron, 2016a). However, early studies identified genes by sequencing candidate genes in loci with frequent deletions and/or duplications in autism.

Bourgeron and colleagues used this approach to identify three early genes with autism: SHANK3, NLGN3 and NLGN4X (Jamain et al., 2003; Durand et al., 2007). NLGN3 and NLGN4X code for postsynaptic, transmembrane proteins called neureligins that contribute to the formation and maintenance of synapse. Following the observations of de novo deletions at Xp22.3, Jamain and colleagues sequenced three neurelin genes, NLGN3, NLGN4X and NLGN4Y, in individuals with autism and identified mutations in two families (Jamain et al., 2003). This was followed by the sequencing of SHANK3, which codes for another postsynaptic protein that binds with neureligins (Durand et al., 2007). SHANK3 is in 22q13, microdeletions which were known to contribute to developmental delay and autistic behaviour. Three of the families sequenced carried mutations in SHANK3 providing evidence for the association of SHANK3 with autism. Subsequent gene sequencing efforts in large cohorts has provided further support for NLGN3 and SHANK3 in autism (Sanders et al., 2015).

Candidate gene approaches have also provided several false positives. Several genes have been investigated in autism using a candidate gene association approach (for a list, see: https://gene.sfari.org/autdb/HG_Home.do). These were typically but not always conducted in relatively small sample sizes, investigating a small number of genetic variants. Warrier, Chee and colleagues conducted a meta-analysis of candidate gene association studies in autism where they reviewed the evidence of 552 genes that have been included in association studies in autism (Warrier et al., 2015a). Common genetic variants in only 27 of these genes had been investigated in three or more independent cohorts, suggesting a scarcity of well-replicated genetic associations for autism. None of the variants included in the meta-analysis were significant in a larger genome-wide association cohort.

Genome-wide association studies of autism and related traits

A few genome-wide association studies have also investigated the genetic architecture of autism with limited success. In contrast to candidate gene association studies, genome-wide association studies typically investigate hundreds of thousands of genetic variants across the genomes and then correct for the total number of independent statistical tests performed. Owing to the number of tests performed, sample sizes have to be several orders of magnitude larger than candidate gene association studies to identify significant associations.

The first study, by Wang et al. (2009), investigated 780 families initially and a second cohort of 1453 autistic individuals and 7070 controls. Meta-analysis of the two cohorts identified one locus at 5p14.1 that was significant ($P<5 \times 10^{-8}$). This intergenic locus was located between two cadherin genes (CDH10 and CDH9) that are involved in diverse neural functions and contain a unique calcium-binding domain. They replicated the locus in two smaller, independent cohorts. Interestingly, this region was also implicated in another genetic association study that used data from 438 autistic families (Ma et al., 2009). SNPs in 5p14.1 were nominally significant ($P<0.05$) although it did not reach genome-wide significance.

Despite these early success, subsequent studies have not been able to replicate association at 5p14.1 at a genome-wide association level, leading to the conclusion that these early studies were statistically underpowered and the effect size inflated due to winner’s curse (or regression to the mean, where the effect size that are most likely to cross the threshold of significance are likely to be inflated when the statistical power is limited).

Four further studies have reported significant association results. In 2009, Weiss, Arking and colleagues (Weiss et al., 2009) used data from multiple different cohorts and meta-analysed results from transmission disequilibrium tests and association studies to identify one SNP that was significant at 5p15.2 between genes SEMA5A and TAS2R1. However, the $P$-value threshold used for significance was not the traditional GWAS threshold of $5 \times 10^{-8}$, but a more liberal threshold of $2.5 \times 10^{-7}$, identified using permutation and after accounting for LD. Another study by Anney and colleagues (Anney et al., 2010) divided participants into four groups along two axes: one
on ethnicity (primary European ancestry vs all ancestry) and one on diagnosis (strict autism vs inclusive spectrum). They identified an intronic SNP in MACROD2 that was associated with strict autism below a genome-wide threshold. A further study conducted a cross-ethnic meta-analysis (European ancestry and Chinese ancestry) to identify variants associated with autism. Meta-analysis identified common variants at the 1p13.2 locus associated with autism. Interestingly, 1p13.2 has been previously been linked to autism in linkage studies (Xia et al., 2014).

Recently, work from the Psychiatric Genomics Consortium using genetic data on more than 16,000 individuals identified an association at 10q24.32 (rs1409313, \( P = 1.05 \times 10^{-8} \)) (The Autism Spectrum Disorders Working Group of The Psychiatric Genomics Consortium 2017).

Despite the general lack of replicably associated loci, there is considerable evidence for a role of common genetic variants, en masse, identified using different SNP heritability measures. Using data from simplex (only one individual with an autism diagnosis) and multiplex families (at least two individuals with an autism diagnosis in the family), Klei et al. (2012) have identified a narrow-sense heritability between 40% and 65% for autism. The heritability was higher for multiplex families (65%) than for simplex families (~40%), consistent with the observation of higher number of de novo loss of function (dnlof) mutations in simplex probands than in the general population (Iossifov et al., 2014; Kosmicki et al., 2017; Sanders et al., 2015).

This estimate of narrow-sense heritability was confirmed by another study that suggested that the majority of the genetic risk for autism is attributable to common variants (Gaugler et al., 2014). The same study identified that dnlof mutations contributed to ~3% of the variance but explained a significant proportion of individual liability. Summing up the contribution from different classes of genetic risk, genetic variants explained approximately about 60% of the total variance (Gaugler et al., 2014). The limited success in identifying significant loci despite the considerable SNP heritability may be due to multiple reasons, including the high polygenicity of the condition and the considerable phenotypic heterogeneity.

Other studies have tried to investigate the genetic architecture of specific subtypes of autism. Two studies have investigated the genetic architecture of Asperger syndrome, a subtype of autism where individuals have average or above average intellectual ability and preserved language. Neither of the studies, however, identified genome-wide significant association (Salyakina et al., 2010; Warrier et al., 2015b). A third study conducted family-based association analysis based on IQ and symptom profiles but did not identify significant differences in heritability nor significant associations, possibly owing to the reduced sample size and, consequently, statistical power (Chaste et al., 2015). As there is considerable phenotypic differences between males and females, another study also investigated sex-stratified analysis but did not find evidence for higher genetic risk for females with autism (Mitra et al., 2016).

A few studies have also investigated the genetic architecture of traits related to autism, given the considerable twin heritability of autistic and related social traits. Two studies have investigated the genetic architecture of social communication in children. Using data from a longitudinal cohort (ALSPAC), St Pourcain and colleagues investigated the SNP heritability of social communication difficulties using the Children’s Communication Checklist (CCC) and the Social-Communication Disorders Checklist (SCDC) (St Pourcain et al., 2013, 2014). Both phenotypes were modestly heritable, with an SNP heritability estimate of the 0.18 for the CCC-derived phenotype and SNP heritability estimates ranging from 0.08 to 0.45 for the SCDC across different ages. Work from our laboratory has also identified significant SNP heritabilities for traits related to autism. Measures of empathy (cognitive and self-report) had significant SNP heritabilities of between 0.05 and 0.12 (Warrier et al., 2016; Warrier et al., 2017).

Development in statistical methods has also allowed for the interrogation of SNP coheritabilities (or genetic correlations) between autism and different phenotypes. Work by Robinson, St Pourcain and colleagues identified a significant and replicable genetic correlation between clinically diagnosed autism and the broader autism phenotype measured using the SCDC (\( r_g \sim 0.30 \)) (Robinson et al., 2016). However, this was dependent on the age at which the SCDC was completed by individuals, with the highest genetic correlation in childhood, which declined with age (St Pourcain et al., 2017). Interestingly, there was no significant genetic correlation between autism and different measures of empathy (Warrier et al., 2016; Warrier et al., 2017).

In parallel, genetic correlation analyses have also investigated the shared genetic architecture between autism and nonsocial traits. The most interesting of these is the consistent positive genetic correlation between autism and different measures of cognition including educational attainment, childhood and adult cognition and number of college years (Bulik-Sullivan et al., 2015; Clarke et al., 2015; Sniekers et al., 2017). Select traits that share a phenotypic and genetic correlation with measures of cognition are also positively genetically correlated with autism such as systemising (an interest in rule-based systems) (Warrier et al., 2016).

This is in contrast with several studies that have identified a significant comorbidity and ID (Lai et al., 2013). There is some epidemiological and genetic evidence of different genetic architecture of autism with versus without ID (Robinson et al., 2014), although a comprehensive discussion is outside the scope of this article. A recent study, however, did not identify a difference in polygenic transmission for autism or educational attainment genetic scores between individuals with autism with versus without ID, providing support for a two-hit model (wherein a dnlof mutation in combination with a background of genetic risk predisposes an individual to autism as opposed to a dnlof mutation alone) (Weiner et al., 2017). Interestingly, a recent study has identified that common genetic variants associated with different signatures of positive evolutionary selection in humans, and this may possibly linked to the underlying pleiotropy with different measures of cognition (Polimanti and Gelernter, 2017).

The current genetic correlation results are limited by the sample size and the effect statistical power of the autism GWAS analysis. Work from the Psychiatric Genomics Consortium (forthcoming) using a larger autism GWAS sample has identified several additional significant correlations including with other psychiatric conditions. See also: Genome-Wide Association Studies.
Copy Number Variation in Autism

Copy number variants (CNVs) are submicroscopic genomic deletions or duplications that are larger than 1000 nucleotides (an arbitrary number). These are frequent in the genome and alter gene dosage (Zarrei et al., 2015). Several de novo and inherited CNVs have been identified in autism. In 2007, Sebat and colleagues published the first investigation of CNVs in autism. Using data from 118 simplex families, 44 multiplex families and 99 control families, they identified a significant excess of de novo CNVs in simplex probands (Sebat et al., 2007). Since then, several studies have investigated the role of CNVs in autism and replicated the initial results (reviewed in (Chung et al., 2014; Geschwind and State, 2015a)). Several well-validated results have emerged. CNVs are found in significantly higher frequency in probands compared to unaffected siblings (approximately 2–3 times more CNVs than siblings) (Sanders et al., 2015). Further, in probands, the CNVs tend to affect a larger number of genes, altering gene dosage of multiple genes. Consistent with the Carter effect, female probands carry more de novo CNVs than male probands (Sanders et al., 2015).

In addition, the number of de novo CNVs is significantly associated with lower IQ, a finding that has been replicated in multiple studies (Leppa et al., 2016; Levy et al., 2011). Overall, de novo CNVs are thought to contribute to small but significant proportion of risk in autism, with 5–15% of individuals with autism carrying de novo CNVs compared to only 1–2% in the general population (Geschwind and State, 2015a). While several CNVs have been identified, it has been challenging to identify CNVs at genome-wide significance due to the difference in lengths and number of genes affected by different CNVs at the same locus. By investigating de novo CNVs in multiple large cohorts, Sanders and colleagues identified six risk loci: 1q21.1, 3q29, 7q11.23, 16p11.2, 15q11.2-13 and 22q11.2 (Sanders et al., 2015).

Insights from Next-generation Sequencing Studies

The sequencing of the human genome in 2004, along with the advent of next-generation sequencing methods, allowed for large-scale hypothesis-free interrogation of the genome in autism using whole exome and, more recently, whole genome sequencing studies of large cohorts. These studies have identified a prominent role for dnlof mutations in autism. Although dnlof mutations have large effect sizes in comparison to common variants, their relative rarity in the population in addition to the genetic heterogeneity in the population makes it difficult to identify genes at a genome-wide significant threshold. Considerable advances have been made in identifying high confidence genes in autism using next-generation sequencing in largely simplex families, and there has been convergence on key findings (De Rubeis et al., 2014; Iossifov et al., 2014; Sanders et al., 2015; Samocha et al., 2014; Neale et al., 2012; O’Roak et al., 2012a,b).

Similar to de novo CNVs, dnlof mutations are enriched in simplex autism (where one child has a diagnosis of autism and the parents and other siblings do not have an autism diagnosis) compared to controls. Female probands are likely to harbour more dnlof mutations than male probands, which is in line with the Carter effect. Similar to de novo CNVs, dnlof mutations are also associated with lower IQ and more severe autistic phenotypes. It is also clear that the number of dnlof mutations increases with paternal age, likely because spermatogonia undergo more active mitosis to produce sperm cells. Finally, these studies have also identified that de novo missense mutations and inherited loss-of-function mutations show smaller effects than dnlof mutations.

Recent efforts have integrated data from multiple sources in order to identify genes that are frequently mutated. Sanders and colleagues (Sanders et al., 2015) identified 65 high confidence genes (false discovery rate <0.1), although it is estimated that between 450 and 1000 such genes may be involved in autism (Geschwind and State, 2015a). Two recent studies have expanded and refined this list of genes using different methods. Yuen and colleagues performed whole-genome sequencing on multiplex autism families and identified 18 additional genes (Yuen et al., 2015; Yuen et al., 2017). In parallel, Kosmicki and colleagues utilised genetic data from a large-scale population resource – the Exome Aggregation Consortium (Lek et al., 2016) – to identify which de novo variants are not observed in the general population (Kosmicki et al., 2017).

Partly due to the families sequenced and partly due to the underlying genetic architecture, many of the genes identified turn out not unique to autism. Indeed, several of these genes are also seen in conditions such as ID and schizophrenia (Geschwind and Flint, 2015), a feature that is also shared by CNVs identified in autism and common genetic variation. This shared pleiotropy among autism, ID and schizophrenia, among other conditions, suggests that a combination of different genetic and environmental effects shapes the disease-specific mechanisms. Investigation of transcriptomic signatures of autism and ID has identified distinct biological networks that contribute to ID and autism (Parikshak et al., 2013).

The use of large-scale data has also allowed for convergence in identifying pathways, particularly synaptic function, chromatic remodelling and downstream targets of FMR1, MECP2 and CHD8 (Pinto et al., 2014; Bourgeron, 2015). Several of the genes identified are associated with postsynaptic density, and many of the mutations are thought to affect synapse formation and plasticity, both during child development and in adulthood. Functional studies using postmortem brain tissue, and animal models, have identified altered synaptic development and pruning (Zoghbi and Bear, 2012; Tang et al., 2014). A few genes identified have also been associated with specific features, leading to subtypes of autism associated with mutations in specific genes.

For example, CHD8 is the most frequently mutated gene in autism. CHD8 encodes for a transcriptional repressor and remodels the chromatin by recruiting histone H1. Phenotypically, CHD8 mutations cause macrocephaly (enlarged head circumference) and increased brain size and persistent gastrointestinal issues, which have been modelled in zebrafish (Bernier et al., 2014). These are all clinical signs in subgroups on the autism spectrum. Individuals with CHD8 mutations also have similar facial dysmorphic features, including widely spaced eyes and a broad nasal tip, alongside cognitive impairments (Bernier et al., 2014).
In parallel, sequencing studies have also identified mutations in CTNNB1, a gene that encodes a protein (beta-catenin) that closely interacts with CHD8. CHD8 negatively regulates (Helsmoortel and in some cases autism (Krumm genetic variation and environment, in modulating the effects of these genes have complete penetrance for autism, suggesting a small but significant role for other factors, including background genetic variation and environment, in modulating the effects of the mutation in these genes.

Transcriptional Dysregulation in Autism

As autism is a neurodevelopmental condition, studies have also sought to investigate alterations in gene expression directly in the developing brain. Parikshak et al. (2013) and Willsey et al. (2013) both used transcriptomic data from developing cortical tissues to identify if genetic risk for autism shows spatiotemporal convergence, using different methods. Willsey and colleagues constructed gene coexpression network using high-confidence autism genes as seed genes and investigated the enrichment for probable autism genes across multiple temporal and spatial windows. They identified enrichment in midfoetal prefrontal and primary motor-somatosensory cortex. Further, by investigating layer-specific gene expression, they were further able to identify enrichment in the cortical innerplate in the midfoetal prefrontal and primary motor-somatosensory cortices.

Parikshak and colleagues used a different approach to investigate convergence of genetic risk in autism in the developing brain. Using whole-genome transcriptome data, they constructed weighted gene coexpression modules agnostic of relationship to candidate genes in autism and followed their expression trajectories across developmental time. They identified three gene coexpression modules that were enriched for candidate genes as seed genes and investigated the enrichment for probable autism genes across multiple temporal and spatial windows. They identified enrichment in midfoetal prefrontal and primary motor-somatosensory cortex. Further, by investigating layer-specific gene expression, they were further able to identify enrichment in the cortical innerplate in the midfoetal prefrontal and primary motor-somatosensory cortices.

A few studies have also investigated gene dysregulation in adult cortical and subcortical tissues, by systematically identifying differentially expressed genes in the autism postmortem brains compared to control postmortem brains (Gupta et al., 2014; Voineagu et al., 2011; Parikshak et al., 2016). These studies have identified several important results. First, cortical gene expression can help to separate the transcriptomes of autistic individuals from population controls. Second, these studies have been able to identify differentially expressed genes (both upregulated and downregulated) in autism cortex compared to control cortex, although analyses of the autism cerebellum in comparison to the control cerebellum have not been forthcoming.

These differences are likely to extend to other cortical and subcortical regions. For example, there is evidence that typical differences in gene expression in the frontal and temporal cortices are altered in autism (Parikshak et al., 2016). Third, there is replicable evidence to suggest that differentially downregulated genes are associated with neuronal and synaptic pathways, whereas upregulated genes are associated with glial (microglia and astrocytes in particular) and immune-related pathways.

Integrative transcriptome analyses have also identified similarities and differences across multiple conditions. A meta-analysis of gene-expression microarray data across cortical transcriptional data sets has revealed overlapping neuropathology between autism and psychiatric conditions such as schizophrenia, bipolar disorder and major depression (Gandal et al., 2016). The correlation between transcriptional dysregulation between these conditions parallels the genetic correlation identified using GWAS data. This shared pathophysiology was replicated using ribonucleic acid (RNA) sequencing data. Another study provided independent evidence for the shared pathophysiology between autism and schizophrenia, by significant and correlation of the transcriptional dysregulation between the two conditions (Ellis et al., 2016).

Conclusions

Autism is an extremely heterogeneous condition. Twin and familial recurrence rates have established significant heritability for the condition and related traits. Rare and common genetic variants, along with copy number variation and transcriptional dysregulation, contribute to risk for the condition. The high genetic and phenotypic heterogeneity makes gene discovery challenging. Interrogating large cohorts have identified some genes and CNVs that contribute to risk for autism, although these represent a small proportion of the genes that are hypothesised to contribute to the condition. Mutations in some genes contribute to specific subtypes in autism with shared clinical and behavioural characteristics.

No common variants have been consistently associated with autism, although it is clear that they contribute, en masse, to a significant proportion of the underlying risk. Bivariate and multivariate analyses have identified considerably pleiotropy with several other conditions, including measures of condition. This pleiotropy is also observed at the level of the cortical transcriptome. There has been considerable effort to develop larger and richer genetic databases of autism, and interrogation of these larger data sets will considerably advance our understanding of the genetics and biology of autism.

Acknowledgements

We thank Helena Brunel and Richard Bethlehem for their comments on the manuscript. Research at the Autism Research Centre is funded by grants from the UK Medical Research Council, the Wellcome Trust, the Autism Research Trust and the Templeton World Charity Foundation, Inc. The opinions expressed in this publication are those of the authors and do not necessarily reflect the views of Templeton World Charity Foundation, Inc. VW is funded by St. John’s College, Cambridge, and Cambridge
Commonwealth Trust. The research was funded and supported by the National Institute for Health Research (NIHR) Collaboration for Leadership in Applied Health Research and Care East of England at Cambridgeshire and Peterborough NHS Foundation Trust. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health.

**Glossary**

**Association study** Studies that systematically test for differences in the frequency of variation between cases and controls or for association between allele frequency and a quantitative trait. Candidate gene association studies investigate variations in specific genes. Genome-wide association studies investigate variations across the genome.

**Common variation** Genetic variation where the minor allele is observed in at least 1% of the population.

**Heritability** Proportion of the total variance in a condition or trait that can be attributable to genetics.

**Rare variation** Genetic variation that is observed in less than 1% of the population.

**References**


The Genetics of Autism


Further Reading

