Chapter 14

The Genetic Architecture of Autism and Related Conditions

Michael Gill, Graham Kenny and Richard Anney

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/54854

1. Introduction

Autism is a behaviourally defined developmental disorder characterised by impairments in social communication, restricted interests and repetitive behaviours [1]. Abnormalities in these three developmental areas tend to cluster together in affected individuals. In DSM-IV, Autism is part of a larger continuum of disorders collectively called Pervasive Developmental Disorders. Autism spectrum disorders (ASD) refer to Autism, Pervasive developmental disorder, not otherwise specified, and Asperger syndrome. All individuals with ASDs have qualitative abnormalities of social development in combination with disorders of communication and/or stereotyped repetitive interests and behaviors. The social skills that develop naturally in typically-developing children do not do so in children with ASD. In addition, there are several behaviors and co-morbid symptoms that relate to each of the three classical impairments. Recent studies have reported rates of co-occurring intellectual disability in the range of 25-50%. Neither developmental delay nor cognitive impairment are required for an ASD diagnosis.

Fombonne and colleagues recently estimated the prevalence of strictly-defined autism at approximately 15-20 per 10,000 people [2]. When the definition of autism is relaxed to include Autism Spectrum Disorders, the prevalence estimated expands to approximately 60 in in 10,000 children [2, 3].

Little is known of the biological basis of ASD and the future development of rational knowledge based treatments will depend on a comprehensive understanding of innate biological predisposition and its interaction with environmental factors. The identification and characterisation of the genetic variation and genes involved in ASD is a route towards this goal. This chapter outlines the various approaches that have been applied to this task, in the context of rapidly evolving technology and human genome resources, and summarises the state of
2. Autism is a heritable disorder

There is unequivocal evidence to support the role of genetic liability as the basis of familiality in the aetiology of ASD. In their seminal work, Folstein and Rutter [4] observed that 4 of 11 (36%) pairs of identical twins were concordant for strictly defined autism – whilst none of 10 pairs of fraternal twins were concordant. In a large follow-up study, which included these original twins, Bailey and colleagues observed a more striking concordance rate in identical twins (60%) with no concordance in the fraternal twins [5]. This highly cited study underlined ASD as the most heritable of the neurodevelopmental and psychiatric disorders with heritability estimates of 91 - 93%. A recent review of over 30 twin studies of ASD [6] noted that the early twin studies, which relied upon a strict diagnosis of autism, show a median identical and fraternal concordance of 76% and 0% respectively. The expansion of these and other studies to include a broader definition of ASD revealed a median concordance rate of 88% and 31%. In both cases, these data reveal that ASD is underpinned by a strong genetic component. More recently, Hallmayer and colleagues [7] using more stringent clinical diagnostic tools, such as the Autism Diagnostic Interview (ADI; [8]) and the Autism Diagnostic Observation Schedule (ADOS; [9]) in a group of 202 twin pairs from the California Twin Registry, identified similar concordance rate of ASD as previously attained; with identical twin boys showing a concordance rate of 77% and fraternal twin boys 31%.

3. Genetic linkage and association

The heritability estimates above calculated from twin studies imply that there is unknown genetic variation in children with ASD that confers risk. Several models may fit the observed family and twin data including, at the extremes, a ‘common disorder-common-variant’ model, with many DNA variants of small effect combining together to create risk; and a multiple rare variant model where ASD is in reality a large collection of different and individually rare disorders. The true picture is only slowly emerging, and as with many other common disorders, common low penetrant and rare higher penetrant DNA variants combine to jointly contribute to risk. Molecular genetic methods have developed exponentially over the last few decades with ever increasing knowledge of the structure and function of the human genome. For disorders with largely unknown aetiology, such as ASD, using human genetic studies to identify risk genes is a gateway to an understanding of underlying biology. Such studies may focus on individuals or families each using different but related genetic strategies to track down risk or causative genetic variation. These methods include genetic linkage, genetic association, and direct examination of structural and copy number variation.

The basis of genetic linkage is the physical co-location of genes or other DNA variation. Such DNA variation is more likely to be transmitted together (or linked) than those that are further
Linkage studies take advantage of this non-random assortment of genetic variation. A linkage study calculates whether a known genetic variant and a disease mutation (represented by the disease trait) are linked and if so, roughly localises the causative mutation. Following the successes of using these approaches in the discovery of multiple loci implicated in Mendelian disorders, researchers were encouraged to apply linkage methodology to more complex traits, such as ASD where Mendelian principles may apply, at least in a proportion of families. However, in ASD, only a few scans have highlighted loci with significant linkage, highlighting loci including chromosome 7q, 2q and 3q.

In 1998, the International Molecular Genetic Study of Autism Consortium (IMGSAC) [10] reported modest evidence for linkage. This included significant linkage arising on chromosome 7q32-q34. Supplemented analyses of IMGSAC families provided additional support for linkage at 7q22, 16p13, and 2q31 [11, 12]. The long arm of chromosome 7 has received particular attention with additional support reported for 7q21 [13], 7q22 [14], 7q31 [15, 16], 7q32 [17] and 7q36 [16].

Few individual families exist where ASD segregates in an obvious Mendelian fashion that is large enough to provide significant evidence for linkage by themselves. Most linkage studies required the assumption that a significant proportion of the families in the sample might be linked to a given locus and few were sufficiently large to accommodate even modest locus heterogeneity. Under a ‘common disorder - common variant’ model, multiply affected families will occur but linkage methods would be considerably underpowered. The effects of DNA variation with low penetrance are more easily identified using a genetic association study design in a sample of cases drawn from a population.

Fine-mapping and candidate gene association studies at implicated regions on 7q have implicated a number of potential susceptibility genes including RELN, MET, CNTNAP2 and EN2. Persico and colleagues [18] studied five DNA variants or polymorphisms across the RELN gene locus, including a GGC repeat variant located close to the RELN gene translation initiator codon. Located at 7q22, RELN encodes an extracellular matrix protein Reelin, which plays a pivotal role in the development of laminar structures including cerebral cortex, cerebellum and hippocampus. Using a genetic association approach, Persico and colleagues identified a nominally significant association with this 5’-UTR GCC-triplet-repeat polymorphism [18]. This finding was further supported by some studies [19-25] but not others [26-31]. MET, located at 7q31, received considerable attention following a high-profile association reported by Campbell and colleagues. The MET gene encodes a protein involved in MET (Mesenchymal epithelial transition factor) receptor tyrosine kinase signalling which has been implicated in brain growth and maturation – offering biological plausibility to its candidature. As with other candidate genes in ASD, the original findings have been supported in some [32], but not other studies [33] [34]. A similar scenario played out for the EN2 homeobox gene located at 7q36 [35, 36]. Arking and colleagues [37] observed an association at CNTNAP2 in the NIHM/AGRE collection. The main association observed was for rs7794745 located in intron 2 of the gene (Discovery $P = 0.00002$, Validation $P = 0.005$). The CNTNAP2 (7q35) gene encodes the contactin-associated protein-like 2 protein, which is a member of the neurexin family and thought to play a role in axonal differentiation and guidance. Li and colleagues also found mild support for CNTNAP2...
where they observed a weak association with a haplotype containing the SNP rs7794745 [38]. More recently, a large GWAS of 2705 families identified a strong association for the SNP rs1718101 in the CNTNAP2 gene with a subset of individuals without intellectual disability and of European ancestry ($P = 7.78 \times 10^{-9}$) [39].

The Paris Autism Research International Sibpair (PARIS) study [15], that supported the findings on 7q, noted additional loci at 2q31-q32, 16p13 and 19p13, that overlapped to some extent with regions identified in the IMGSAC analyses. Fine mapping of the 2q31-q32 led to additional evidence implicating the genes SLC25A12, STK39 and ITGA4 as putative risk genes for ASD. As with most candidate genes examined in small samples, these associations were validated by some [40-42] but not all studies [43-46]. The positional candidature of the 2q31-q32 was further supported by a case study of a young Irish male with high-functioning autism with a complex translocation traversing chromosome 2q (46,XY,t(9;2)(q31.1;q32.2q31.3)) [47]. Fine mapping and mutation analysis identified an association with a polymorphism within the splice donor sequence of exon 16 of the ITGA4 gene and ASD (rs12690517; $P = 0.008$) [48].

In a study of Finnish families, Auranen and colleagues [49, 50], who included individuals with autism, infantile autism, Asperger syndrome (AS) and developmental dysphasia reported a significant linkage to 3q25-q27. This was supported by suggestive evidence at 3q25-q27 in a study of AS also in the Finnish population [51] as well as linkage at 3q25-q27 in a single large extended Utah pedigree of Northern European ancestry [52].

Interestingly, the advent of large collaborative studies by the Autism Genome Research Exchange (AGRE), the Autism Genome Project (AGP) and the National Institute of Mental Health (NIMH) has not yielded stronger evidence in favour of specific loci. Liu and colleagues using data from 110 multiplex families from the AGRE collection [16] observed only suggestive linkage on chromosomes 5p13, Xq26-qtel, and 19q12 alongside modest support for previously reported linkage on 7q (7q31 and 7q36) and 16p13. A follow-up analysis including 235 additional AGRE multiplex families was again limited to only suggestive loci at chromosomes 17q11, 5p13, 11p11-p13, 4q21-q22 and 8q24 [53]. In a much larger collection of 1168 multiply-affected families from the AGP (including families previously included in the AGRE, CPEA, IMGSAC, PARIS and Seaver linkage studies), Szatmari and colleagues identified suggestive linkage to chromosome 11p12-p13 and a large region on chromosome 15q23-q25 [14]. Of the regions that were featured prominently in previous linkage analyses, there was only modest support for previously highlighted linkage regions on chromosome 2q31 (female autism-probands) and 7q22 (male ASD-probands) from families of European ancestry [14]. A similarly large linkage study of 1031 families, including 1553 affected offspring, Weiss and colleagues identified suggestive linkage at 6p27 and significant linkage at 20p13 [54].

Early association studies examined whether specific variation in genes was associated with the disease and focused on candidate loci identified from linkage and cytogenetic studies (positional candidate genes) as well as in genes within biological processes that we perceived as having a role in ASD. In the mid-1990’s Risch and Merikangas [55] demonstrated that where genetic variants have only small effect on risk the association study is a more powerful ap-
proach than linkage to identify genetic risk. However, the transition from candidate gene to genome-wide association studies was not realised until technological advances firstly enriched the maps of common variation across the genome and secondly enabled the interrogation of these variants en masse as part of genome-wide SNP arrays.

Wang and colleagues [56] performed one of the first GWAS on individuals in European ancestry individuals from the AGRE collection, the Autism Case Control Collection and unaffected controls from the Children’s Hospital of Philadelphia control collection. Neither family-based nor case-control analysis alone yielded genome-wide significant findings. However, in a combined analyses the authors identified genome-wide significant association on chromosome 5p14.1 (rs4307059; $P = 3.4 \times 10^{-8}$) and a number of additional association signals on chromosome 13q33.3, 14q21.1 and Xp22.32. The 5p14.1 association was validated in the Collaborative Autism Project (CAP) and Centre for Autism Research and Treatment (CART) study. The authors found a modest to strong replication of the association signal on chromosome 5p14.1. In a reciprocal study with the CAP and CART study as the discovery sample followed by validation using the AGRE dataset was published in parallel by Ma and colleagues [57]. The authors examined approximately 500k SNPs, more than in the Wang report and albeit not genome-wide significant, they retained the association signal on chromosome 5p14.1.

A second independent GWAS was reported by Weiss and colleagues [54]. In the initial scan the authors did not find any GW-significant associations. However, as with the previous GWAS, additional supplementation of their family-based studies with a case-control set derived from 90 probands without parental data garnered some additional signal for the top hits. A replication consortium of greater than 2000 trios was genotyped for 45 SNPs across all of the top associated regions. The only marker that showed evidence of replication resides on the short arm of chromosome 5 at 5p15. Although, like Ma and colleagues, this report has considerable overlap with the AGRE families reported by Wang and colleagues, the authors did not observe strong association at 5p14.1. The chromosome 5p association lies in close proximity to TAS2R1. The TAS2R1 gene encodes a G-protein coupled receptor that is involved in bitter taste recognition. The authors highlight a more biologically plausible ASD candidate gene approximately 80 Kb telomeric, SEMA5A. SEMA5A encodes a gene important in axonal guidance and has been shown to be down regulated in the occipital lobe cortex, lymphoblast cell lines and lymphocytes from individuals with autism.

Two additional GWAS followed by the Autism Genome Project (AGP) [39, 58]. In the first report of 1369 families, Anney and colleagues identified a single GW-significant finding on chromosome 20 at position 20p12 within the MACROD2 (MACRO-domain containing 2) gene locus (rs4141463; $P = 2.1 \times 10^{-6}$; OR = 0.56). Weak statistical support was observed for MACROD2 in an AGRE validation sample, albeit showing the same direction of effect for the risk allele. In a follow-up study using an additional 1301 families the authors showed little if any signal of association (rs4141463; $P = 0.206$, OR=0.91). In a combined analyses the association at MACROD2 was less compelling (rs4141463; $P = 1.2 \times 10^{-6}$; OR=0.77). The role of MACROD2 is largely unknown although structural variation in this gene and specifically the region harbouring the ASD association signal have been implicated in schizophrenia [59, 60] and epilepsy [61]. However, this same region has been identified as a hotspot for deletions in the genome...
identified as a region of large rare deletions [62]. MACROD-like proteins are highly conserved across evolutionary time, which may indicate an essential role. The MACRO-domain is an ADP-ribose binding module [63] and has been implicated in the ADP-ribosylation of proteins, an important post-translational modification that occurs in a variety of biological processes such as DNA repair [64], heterochromatin formation, histone modification and sirtuin biology [65-67] as well as long-term memory formation [68]. The association signal observed at rs4141463, albeit tagged to the MACROD2 gene, resides in an intronic region near an intragenic non-protein-coding-RNA, NCRNA00186 (MACROD2-AS1). Two MACROD2-AS1 transcripts have been reported of 673bp and 1230bp in length, located on the reverse strand between exon 5 and 6 of MACROD2. Anti-sense RNAs typically interact with mRNA, resulting in transcriptional or post-transcriptional effects and have been linked to brain development and plasticity [69]. However, unlike MSNP1AS described for the 5p14.1 association observed by Wang and colleagues, a function for this non-coding RNA has not yet been reported.

The strongest association signal observed by the AGP combined analyses was rs1718101 ($P = 7.8 \times 10^{-9}$; OR=2.13 (1.63-2.80)), a SNP within the CNTNAP2 gene which was previously implicated in ASD through linkage analyses. This association was observed in a secondary analysis which was restricted to ASD individuals of European ancestry with a higher IQ. Anney and colleagues suggest that with the current data “few if any common variants have an impact on risk exceeding (an effect size of) 1.2 (or below its inverse).” In an attempt to seek evidence for or against common variants having an impact on risk, the authors constructed an allele-score. The allele-score method, as previously described by Purcell and colleagues [70], calculates a score for each individual based on number of risk associated alleles that the individual possesses. This score is then used to either calculate the predictive value of the score between cases and controls or estimate the amount of variance that this score predicts for the disease. Allele-scores derived from the transmission of common alleles from the families described in the AGP stage 1 GWAS [58] could significantly predict case-status in the independent Stage 2 sample. The authors concluded however that despite the limited findings for individual loci from GWAS studies to date, en masse the top results exert a detectable impact suggesting that as the sample sizes increase, additional significant loci will emerge. Putting together samples of the size seen in successful GWAS studies in other disorders (n>20,000) is challenging for a disorder with a complex phenotype such as ASD.

4. Structural variation: Chromosomal abnormalities in autism

Autism has been frequently associated with chromosome abnormalities, such as deletions, duplications, inversions, or translocations; with abnormalities on the long arm of Chromosome 15 and with numerical and structural abnormalities of the sex chromosomes. Reddy in a survey of chromosomal abnormality in autism found one in 14 of 421 individuals (3.33%) [71]. These fourteen cases broke down into 4 supernumerary chromosome markers, 4 deletions, 3 inversions and 3 duplications. Several reviews have since confirmed that such abnormalities can be identified in ~3-5% of these patients [72, 73]. The regions commonly reported include 2q37,
5p14, 5p15, multiple locations on chromosome 7, 11q25, 15q11-q13, 16q22.3, 17p11.2, 18q21.1, 18q23, 22q11.2, 22q13.3 and Xp22.2–p22.3.

Many studies have converged on particular chromosomal abnormalities in autism, the most common of which are maternally inherited duplications at 15q11-13. These duplications are found in as many as 1–3% of patients diagnosed with autism.

Another relatively common chromosomal variant is the 22q11.2 microdeletion syndrome, or velocardiofacial syndrome (VCFS) that occurs in ~1/4000 live births. This syndrome is also identified in the context of learning disability or in schizophrenia and has a complex phenotypic expression affecting multiple organs. The physical features include a typical facial appearance, (long face, narrow palpebral fissures, flattened malar eminences, prominent nose and small mouth); anatomical and/or functional abnormalities of the palatal shelves such as cleft palate and velopharyngeal insufficiency; lymphoid tissue hypoplasia and heart defects. A wide range of childhood onset developmental symptoms and disorders are described in association with the VCFS mutation [74] including attention deficit hyperactivity disorder (ADHD), oppositional defiant disorder, phobias, anxiety, obsessive compulsive disorder and autism spectrum disorders.

5. Copy number variation

The human genome has both sequence and structural variation, the majority of which has no functional consequences and is unrelated to any disorder. Structural variation may be balanced, such as is the case for inversions and balanced translocations or may alter DNA copy number. The later, referred to as copy number variants (CNVs) extend from the duplication or deletion of a single base pair to whole chromosome abnormalities. The term CNV is generally used to indicate the larger changes. Initially, large CNVs were identified with classical chromosomal staining and light microscopy. As many of these abnormalities were initially identified in sub-telomeric regions, there was interest in knowing if such deletions, duplications and other re-arrangements might occur throughout the genome. This hypothesis was confirmed with the development of high resolution array based comparative genomic hybridization (aCGH) [75].

In the arrays, comparative hybridization is performed with DNA immobilized on a platform such as a glass slide. Initially the DNA arrays consisted of human DNA cloned into bacterial artificial chromosomes (BAC arrays) representing the human genome at approximately 1Mb intervals. The present arrays consist of 25-base pair oligonucleotides (probes). In the last decade the resolution of the arrays have improved to the extent that the number of smaller CNVs identified in the genome that were previously invisible to microscopy has increased enormously. Deletions and duplications at least as small as 10kb are now known to occur throughout the genome.

Interestingly, some CNVs appear to differ from single nucleotide polymorphisms (SNPs) in terms of locus specific mutation rates. Rates for genomic re-arrangements range from $10^{-4}$ to
10^{-5}, a rate that is considerably more frequent than point mutations [76]. This high mutation rate, coupled with reduced fecundity in some carriers, and the fact that comparatively, they affect a larger proportion of the genome, make CNVs a potentially important source of new and recent mutation in neurodevelopmental disorders.

6. CNVs in autism

As discussed above, autism is a phenotypic feature of many genomic disorders. Betancur [77], in her review, lists 103 disease genes and 44 genomic loci where autism or autistic like behaviors have been described. ASD is diagnosed in ~30% of males with Fragile X syndrome and in reverse, Fragile X mutations are found in as many as 7–8% of individuals with ASD [78]. Similarly, mutations in MECP2, the Rett Syndrome gene, have been found among cases of autism that do not have the classical Rett phenotype and autism patients have an increased risk for neurofibromatosis and other rare monogenic diseases like tuberous sclerosis and Joubert’s Syndrome, again inversely patients with these disorders have an increased risk for having autism [79, 80].

The genes and loci listed in the Betancur study are all causally implicated in learning disability (LD), indicating that these two neurodevelopmental disorders share some genetic risk factors. Early use of aCGH in non-syndromic autism suggested the method had promise in detecting hitherto unrecognized CNVs. For example Jacquemont et al [81] identified 6 deletions and 2 duplications in 29 patients presenting with syndromic ASD where previous high resolution karyotyping was reported as normal. Another study showing the potential for CNV analysis was a large linkage study [14] using a 10k SNP array where intensity data was used to determine copy number. The authors highlight some individual findings including a family with two sisters with ASD, both of whom had a ~300kb deletion on Ch. 2p16 that included the coding exons of the neurexin 1 gene (NRXN1). A second finding was a recurrent 1.1Mb duplication at Ch. 1q2.1 in four affected individuals from three families, a third was a ~900kb de novo duplication at 17p12 in an affected sib-pair, and with the same region appearing as a maternally inherited deletion in two male siblings, with a paternally inherited deletion in a further female. Duplications in this region cause Charcot-Marie-Tooth 1A (CMT1A) and hereditary neuropathy with pressure palsies when deleted, and overlapping deletions are seen in Smith-Magenis syndrome that includes autism symptoms in many cases [82].

A key development was the report of de novo copy number variants in autism [83] using aCGH. These authors showed that individually rare CNVs, and in particular ones that affect neurodevelopmental genes, were enriched in cases. They further suggested that the rate of de novo CNVs differed between simplex cases, where they occurred in 10% of families in the sample, and familial cases where they occurred in 3% of families, suggesting that sporadic and familial cases of ASD might have different underlying genetic mechanisms, although not all studies since then have found this distinction. Several studies of CNVs in large autism case and family series have followed. Marshall et al. [84] examined 427 ASD families using a 500k
SNP array and karyotyping by standard clinical diagnostic method. A de novo rate of 7.1% and 2.0% in simplex and multiplex families respectively was observed, supporting the previous findings [83]. Families occasionally showed more than a single de-novo event where both may combine to produce risk. A further set of loci were identified in two or more unrelated families, increasing the evidence supporting a pathogenic role. As with the LD literature, at some loci, both deletions and duplication were found suggesting a more complex mechanism than simple over or under-expression of gene products. Of the 196 inherited CNVs confirmed experimentally, 90 were of maternal and 106 of paternal origin. The authors list numerous potential ASD candidate genes where a structural change was either de-novo, found in two or more unrelated ASD cases, or, in the case of the X-chromosome, transmitted from an unaffected mother. Given their rarity, very few individual CNVs in this study provided statistical evidence to support their role in autism. For example, 4 CNVs from 427 cases were found at the DPP6 – a subunit that affects the function of Kv4.2 channels at the same site of expression as SHANK3 and NLGN gene products. Only one similar CNV was found in 1652 controls (Fisher’s exact test p = 0.016). In keeping with previous cytogenetic findings and the emerging overlap in disorders involved, CNVs were found in ASD cases that involved known loci or genes in disorders such as Waardenburg Type IIa, Speech and language disorder, learning disability and VCFS. A further study [85] found a total of 51 CNVs in 46 cases and not in the controls. 42 of these were familial and 9 de novo with recurrence in two or more cases at three loci. In total, case specific CNVs were found in 11.6% of cases, although, in keeping with the Marshall et al [84] study, none were individually associated with case status with the majority being observed in only a single case. Pinto et al [86] compared CNVs in 996 ASD cases of European ancestry to 1,287 matched controls, using the Illumina 1M SNP array. Cases were found to carry a higher global burden of rare, genic CNVs, especially so for loci previously implicated in either ASD and/or intellectual disability. Nearly 6% of the cases had de-novo mutations with some having two or more events. Novel candidate genes were identified that were de-novo in cases and not controls, including SHANK2, SYNGAP1 and DLGAP2. In keeping with previous studies only one novel CNV (maternally inherited X-linked deletions at PTCHD1) occurred statistically more frequently in cases compared to controls (7 vs 0). PTCHD1 involvement in autism and LD was further extended and confirmed in a focused examination of the PTCHD1 locus in cohorts of autism and LD cases and extending the study of CNVs to sequence data, identifying additional maternally inherited missense mutations in 8 probands not seen in controls [87]. In the Pinto et al study[86], certain gene sets were found to be enriched for case deletions but not duplications. These included sets involved in cell and neuronal development, projection, motility and proliferation; GTPase/Ras signaling known to be involved in regulating dendrite and spine plasticity; and kinase activity/regulation. There was additional overlap with gene sets thought to be involved in LD including microtubule cytoskeleton, glycosylation and CNS development/adhesion. More recently, Salyakina and colleagues [88] have shown the value of extended multiply affected families in a CNV study of 42 families. They found 5 deletions and 7 duplications that co-segregated with ASD, two overlapping with known autism CNVs on 7p21.3 and 15q24.1 and two near regions on 3p26.3 and 12q24.32 previously associated with schizophrenia.
As the resolution of the probe arrays improves, smaller CNVs will be detected, and the boundaries of previously identified CNVs will become more refined [89]. Nord et al. [90], examined genomic DNA of 41 children with autism and 367 healthy controls for rare CNVs using a very high-resolution aCGH platform. They found that cases were more likely than controls to have CNVs as small as ~10 kb, likely to affect genes involved in transcription, nervous system development, and receptor activity. They found that expression of CNTNAP2, ZNF214, PRODH and ARID1B genes affected by CNVs were decreased in probands compared with controls suggesting reduced expression as a potentially aetiological factor during development.

Larger samples, particularly those based on families, will also enable the improved estimation of the overall effects of de-novo mutations and the assessment of rare recurrent events as disease associated mutations [91]. These authors studied 1124 autism families containing probands, unaffected parents and an unaffected sibling using the Illumina 1M SNP array. In a related paper [92] they were able to confirm de novo CNVs identified using the SNP array with those detected using a Nimblegen 2.1M aCGH platform. A combined total of 58 rare de novo CNVs were identified across the two studies with each array type identifying 95% of the total. However, the sensitivity for smaller CNVs was low for both arrays. Overall, the burden of rare de novo CNVs in the Sanders et al [91] study was greater in probands than in siblings for total number, size and gene content. Using the rate in siblings as a control to evaluate findings in the cases, there was strong individual statistical support for recurrent de novo duplications at 7q11.23, the locus at which deletions cause Williams-Beuren syndrome; deletions at 16p11.2 and duplications at 16p11.2. In addition the authors observed 8 loci at which rare transmitted CNVs, present only in probands, overlapped with one of the 51 regions in probands containing one or more rare de novo CNVs. However, the rare transmitted CNVs were not more likely to be in cases than in unaffected siblings, even when subdivided into genic, exonic, brain-expressed or previously identified as ASD related. This suggested that the excess burden in their sample was due to rare de novo events, although when the gene sets were applied to gene pathway analysis, more pathways showed enrichment in the case set compared to the sibling set. To date, the number of definitive replicated findings for ASD from all studies has been small, with the data suggesting an extreme heterogeneity model with no single risk variant occurring in more than 1% of cases.

7. Mutation rates and models of risk in autism

Given the replicated finding that de-novo mutations are more frequent in simplex cases compared to familial cases, Zhao et al [93] have suggested a model of autism risk in which families fall into two groups; those in which the overall risk for autism is low, representing the majority of families, and those in which the risk is higher due to a disease mutation with a dominant mode of transmission with greater penetrance in males compared to females. Under this model, sporadic cases of autism occur in low-risk families due to a de novo mutation of relatively high penetrance, whereas familial autism occurs due to the inheritance of an existing mutation from a clinically unaffected or asymptomatic parent. In another model, Girirajan et al. [94] proposed the necessity in some families for a second mutation to lead to severe neurodeve-
lopmental disorder. In this study, individuals with childhood developmental delay are enriched approximately fourfold for a rare 520-kb 16p12 deletion. In nearly all cases examined (22/23), the deletion was inherited. Thus, 16p12 deletions appear to be an example of inherited predisposition to neurodevelopmental disorder with dominant transmission. However, these individuals were more likely to carry a second large (>500 kb) CNV compared to matched controls, and clinical features of those with a second large CNV were typically more severe than those with the 16p12 deletion alone. Itsara et al. [95] suggest that multiply affected autism pedigrees segregate an existing inherited mutation of low penetrance which by itself is rarely sufficient to cause disease. Secondary mutations, such as de-novo mutations are required to manifest as disorder. Whether or not these second hits are disease specific remains to be examined. The authors propose that the excess of de novo CNVs among cases may be due to a depletion of second-hits in the unaffected sibling due to the initial low penetrant mutation segregating in the family. The abundance of inherited low penetrance mutations and the high rate of de novo CNVs in the population enable multiple cases to appear within families with apparently unusual patterns of inheritance.

8. Next-generation sequencing studies

There is an increasing literature describing the occurrence of single nucleotide mutation in ASD. Unlike the common polymorphisms described earlier, these variants are often rare or private to an individual and considered to have a large impact on the gene’s function. Early studies have focused on candidate genes with strong a priori evidence to suggest causation, such as the Neuroligin and Shank encoding genes. Sanger sequencing of the exons of families with ASD have identified many putatively causative mutation in these synaptic proteins [96, 97]. As with the transition from candidate gene association studies to GWAS, the advent of more sophisticated high-throughput next-generation sequencing approaches have enabled many genes to be screened in a single experiment. This has extended to the systematic scanning of the whole exome of individuals with ASD to identify putative risk mutations.

In April 2012, three key manuscripts were published in Nature describing a considerable exome sequencing effort of 622 family trios and 250 unaffected siblings [98-100]. In these studies the focus was on de novo mutation in the exome of individuals with ASD. De novo mutations are those changes that occurred in the child which are not inherited from the parent. Given the natural mutation rate, the number of families studied and the number of genes under investigation it was estimated that three independent mutations would provide strong evidence implicating a gene in ASD, whilst two independent mutations would be suggestive but not definitive. Unfortunately in these studies no gene was burdened with three independent mutations, whilst several revealed two independent de novo mutations. Follow-up analyses across the studies of the two-hit genes in a cohort of over 2600 individuals with ASD and over 1600 typically developing controls identified additional non-synonymous mutations in KATNAL2 (katanin p60 subunit A-like 2), CHD8 (Chromodomain-helicase-DNA-binding protein 8), GRIN2B (Glutamate [NMDA] receptor subunit epsilon-2), LAMC3 (Laminin subunit gamma-3), SCN1A (sodium channel, voltage-gated, type I, alpha subunit) and SCN2A (sodium
channel, voltage-gated, type II, alpha subunit). Despite the early caution regarding the number of de novo events observed per gene, the burgeoning evidence from focused re-sequencing of these suggests that other families with ASD have damaging loss-of-function mutation in these genes.

9. Implications of genomic studies for clinical practice

Genetic findings to date in autism include known genetic syndromes such as Rett and Fragile X; chromosomal structural variation such as duplications on ch15 and deletions on ch22 and increasing numbers of CNVs, many de novo, in a small proportion of cases, especially those with co-morbid developmental abnormalities. Common variants are only slowly emerging from genome-wide studies but sample sizes are still relatively modest compared to other disorders. The diagnosis remains as a clinical one, but genetic testing has diagnostic, prognostic and family planning implications, and in the coming years may also be of importance with respect to potential treatments for specific conditions such as Rett and Fragile X. Therefore, a clinical genetics evaluation should be considered in ASD children in order to identify syndromic forms of autism, identify familial cases, and drive diagnostic testing. Genetic testing using aCGH has been recommended by the American Academy of Pediatrics [101] and the American College of Medical Genetics [102, 103], particularly to evaluate aetiological heterogeneity and identify syndromes. For the vast majority of cases with an identified genetic cause, there is no specific benefit in terms of treatment, but the diagnosis itself may be of value to parents seeking an explanation for the symptoms they see in their child. Acceptance of the disorder may be easier in the presence of a known cause as echoes of social causation and blame can still be heard. Family investigation may also be warranted, particularly where the proband has an identified genetic causation as recurrence risk in siblings may have implications for family planning. If a genetic cause of pathogenic significance is identified, the recurrence risk for sibs may be established according to the particular genetic diagnosis. In the majority of cases, no genetic alteration is found, and an empirical risk for siblings of 10-20% can be provided. Establishing a clear de novo origin for a given mutation may enable a lower risk estimate in siblings although there are caveats relating to the exact origin of the mutation. Finally, and re-stating the main goal of genetic research in ASD, a better understanding of underlying biology may lead to novel and unexpected treatment possibilities. Early signs of this are emerging in human trials for treatments aimed at the specific mutations found in Rett and Fragile X syndromes.

Author details

Michael Gill, Graham Kenny and Richard Anney

Department of Psychiatry, School of Medicine, Trinity College Dublin, Ireland
References


[12] International Molecular Genetic Study of Autism Consortium (IMGSAC), Monaco AP. A genomewide screen for autism: strong evidence for linkage to chromosomes 2q, 7q,


