The role of rare structural variants in the genetics of autism spectrum disorders

M. Kusenda^{a,b} J. Sebat^a

^aCold Spring Harbor Laboratory, Cold Spring Harbor, NY

^bGraduate Program in Genetics, State University of New York at Stony Brook, NY (USA)

Accepted in revised form for publication by H. Kehrer-Sawatzki and D.N. Cooper, 10 September 2008.

Abstract. Autism is a neurodevelopmental disorder characterized by impaired social interaction and communication and restricted interests and behaviors. Despite high estimates of heritability, genetic causes of ASD have long been elusive, due in part to a high degree of genetic and phenotypic heterogeneity (Bailey et al., 1995). Recently, important advances have been made in the genetics of ASD with the use of new technologies for the direct detection of copy number variation (CNV) in the human genome. CNV studies have revealed that *de novo* deletions and duplications, typically less than 1 Mb in size, are strongly associated with ASD, suggesting that spontaneous structural mutations

play a more important role in the etiology of disease than was previously recognized. Rare mutations have been identified at many different locations in the genome, and multiple ‘hot spots’ have been identified where identical rearrangements recur with high frequency. These findings are consistent with the hypothesis that autism, like mental retardation, is caused by a large number of individually rare mutations. These studies serve as a model for how other emerging technologies for mutation detection (e.g. next generation sequencing platforms) could be used to further elucidate the role of rare sequence changes in ASD.

Copyright © 2009 S. Karger AG, Basel

Autism is a class of developmental disorders characterized by impairments in social interaction and communication, and restricted or repetitive behaviors. Because of the substantial clinical heterogeneity of this disorder (Baron-Cohen and Belmonte, 2005), it has become commonly referred as ‘the autisms’ or Autism Spectrum Disorders (ASDs) (Abrahams and Geschwind, 2008). ASD is now recognized as a common disorder, with an estimated population prevalence of 1/150 (Autism and Developmental Disabilities Monitoring Network Surveillance Year 2002 Principal Investigators; Centers for Disease Control and Prevention, 2007); and ASD is four times more prevalent in boys than in girls.

Evidence from twin studies strongly supports autism as a genetic disorder (Abrahams and Geschwind, 2008). With concordance rates for monozygotic twins of 70–90%, and 10% for dizygotic twins, ASD is the most highly heritable of all neuropsychiatric disorders (Freitag, 2006). Early progress in identifying autism susceptibility genes was made through the identification of rare ‘monogenic’ forms of autism, such as Rett syndrome (Amir et al., 1999), Fragile-X syndrome (Verkerk et al., 1991; Yu et al., 1991), tuberous sclerosis (European Chromosome 16 Tuberous Sclerosis Consortium, 1993) and neurofibromatosis (Barker et al., 1987), using classical genetic approaches. Additional autism loci have been identified by candidate gene-based approaches, such as Neurologins 3 and 4 (Jamain et al., 2003; Laumonnier et al., 2004) and Shank3 (Durand et al., 2007). Further evidence for rare mutations in the etiology of ASD comes from cytogenetic studies, where chromosomal abnormalities are consistently reported in 5–7% of patients (Vorstman et al., 2006). The relevance of such rare mutations to the broader patient population was unclear initially, because the individual loci involved each only account for a small fraction of cases. Mutations like those above were gen-

My laboratory is supported by funding from the National Institutes of Health (MH076431-04 and HG004222-02), Simons Foundation and Stanley Foundation.

Request reprints from Jonathan Sebat
Cold Spring Harbor Laboratory, 1 Bungtown Rd.
Cold Spring Harbor, NY 11724 (USA)
telephone: +1 516 422 4198; e-mail: sebat@cshl.edu

erally regarded as uncommon causes of disease with little relevance to the broader patient population.

The prevailing hypothesis for the genetic basis of autism and other neuropsychiatric disorders has thus been the 'common gene/common disease' hypothesis which posits that disease results due to the additive or multiplicative effects of multiple genetic and environmental factors, with individual genes accounting only for a small increase in the risk in an individual patient (Lohmueller, 2003). Consequently, considerable effort has been devoted to the approaches most suitable to test this hypothesis, which are primarily linkage analysis and more recently genome-wide association studies (GWAS). Despite rigorous effort, definitive identification of autism genes by this approach remains elusive. Results from early linkage studies of ASD found evidence for linkage at many different locations throughout the genome (Risch et al., 1999). Several subsequent linkage scans produced similar results (Alarcon et al., 2002; Auranen et al., 2002; IMGSAC, 2001; Liu et al., 2001; Szatmari et al., 2007), but no individual loci identified therein were widely replicated. These findings were consistent with the model that autism is a complex disorder involving a large number of different genes.

Subsequently, with the advent of high-resolution oligonucleotide microarrays, it became possible to test the alternative hypothesis that ASD is caused by rare mutations, consisting in part of CNVs. Microarray comparative genomic hybridization (CGH) provides a means to screen genome-wide for structural changes not detectable using cytogenetic methods. These approaches offered a novel way to perform unbiased mutation screens of the genome in disease studies (Pollack and Iyer, 2002; Shaw-Smith et al., 2004). A CNV scan of the genome, like any classical genetic approach, is done in an unbiased fashion, and risk factors are identified based on the association of genetic markers with case status. However, the CNV-based approach is unique in that many novel rare variants are detected as well as common polymorphisms. This provides an opportunity to look for disease association in new ways. In addition to testing the association of variants individually with disease (single-marker association), one can simultaneously look for multiple rare mutations 'piling up' in the same region. Furthermore, one can examine how the aggregate of rare variants detected in patients differs quantitatively or qualitatively from variation in controls (Sebat et al., 2007; Stone et al., 2008; Walsh et al., 2008; Xu et al., 2008). This approach serves to identify candidate genes that can be further explored through targeted genetic screening as well as molecular studies of gene function.

Early success in CNV studies of ASD

Several studies have recently been published examining the role of structural variants in ASD (Jacquemont et al., 2006; Sebat et al., 2007; Szatmari et al., 2007; Kumar et al., 2008; Marshall et al., 2008; Weiss et al., 2008). Following the clinical cytogenetics model that chromosomal abnormali-

ties are most likely to occur in cases with the most pronounced developmental impairments (Vorstman et al., 2006), Jacquemont et al. (2006) examined a series of 29 patients with 'syndromic' ASD, i.e. cases with a visible growth abnormality or malformation using a 1 Mb-resolution microarray. Seven rare deletions and three rare duplications were detected, ranging from 1.4 to 16 Mb. Seven of these events were de novo, amounting to a de novo rate of 24% in syndromic cases. Two larger studies examined the role of CNVs in idiopathic autism (Sebat et al., 2007; Szatmari et al., 2007; i.e. using patient samples that were not enriched in individuals with multiple congenital anomalies). Szatmari et al. (2007) examined copy number variation in a large sample of concordant sib-pair (multiplex) families analyzed using a 10K SNP platform. A statistical association of CNVs with ASD was not reported. However, some individual rare CNVs were detected in multiple families, and several de novo CNVs were detected including a deletion of a compelling gene candidate, *neurexin-1*, which occurred independently in two affected siblings from the same family. The role of de novo structural variants in idiopathic ASD was specifically addressed in a study by Sebat et al. (2007), which examined the rate of de novo CNVs in patients as compared to healthy controls, and in sporadic cases (simplex families) as compared to patients that have an affected first-degree relative (multiplex families). Using an 85K resolution array, de novo CNVs were detected in 10% of simplex cases, a significantly higher rate than in controls, which was 1% ($P = 0.0005$). By contrast, de novo events occurred in 2% of cases from multiplex families, a modest increase compared to controls. These findings suggested that spontaneous submicroscopic lesions, contribute to disease in a significant fraction of sporadic cases, and to a much lesser extent in multiplex families. The ratio of males to females observed among patients with de novo CNVs is markedly reduced, <2:1 compared to overall patient sample, which is ~4:1, suggesting that these mutations had increased penetrance, contributing to disease more equally in males and females. The vast majority of de novo events were deletions, consistent with deletions being more likely to have deleterious developmental effects than duplications. Recurrent de novo mutations were observed at some loci, but the majority of de novo mutations were unique to individual families. These findings were consistent with the presence of many different autism genes in the genome. Marshall et al. (2008) reported similar findings using an Affymetrix 500K platform in a sample of 427 families, where de novo CNVs were detected in 7% of simplex cases and 2% of multiplex cases. Consistent with our findings, the majority of de novo CNVs detected by Marshall et al. were deletions, and the ratio of males to females among patients was <2:1.

Further studies have begun to reveal which of these novel CNVs occur most frequently. Among the most frequent of the de novo mutations recently discovered is a ~500 kb microdeletion of 16p11.2. First reported in individual patients with congenital malformations (Ghebranious et al., 2007) and autism (Sebat et al., 2007), this mutation has now been shown to occur in ~1% of ASD and other pervasive

Table 1. Recurrent de novo CNVs detected in ASD. Recurrent mutations detected by microarray CGH were compiled from studies by Christian et al. (2008), Jaquemont et al. (2006), Marshall et al. (2008), Mefford et al. (2008), Sebat et al. (2007), and Szatmari et al. (2007). CNVs reported in each study were compared to identify mutations that occurred in multiple unrelated individuals (within and across studies), where a CNV was observed in at least one instance as a de novo event. The minimum overlapping region was determined (genomic coordinates are from May 2004 Build 35). For rearrangements that appear to be mediated by flanking segmental duplications, the boundaries were extended to these regions. References to the studies reporting the variation are designated by a letter corresponding to (A) Christian et al. (2008), (B) Jaquemont et al. (2006), (C) Marshall et al. (2008), (D) Mefford et al. (2008), (E) Sebat et al. (2007), and (F) Szatmari et al. (2007). The number of genes involved was determined based on the minimum overlapping region, including only validated genes found in the UCSC genome Browser (Build 35). Flanking regions, 500 kb on either side, were examined for the presence of paralogous segmental duplications, and these are listed.

Minimum overlapping region	Length	Copy number change	Number/name of genes	Flanking segmental duplications	Frequency	References
chr1:143,650,000–145,000,000	1,350,000	loss	9	yes	1 de novo, 1 inherited	D
chr1:143,650,000–145,000,000	1,350,000	gain	9	yes	1 de novo, 3 inherited	D
chr2:50,430,000–50,785,000	355,000	loss	<i>NRXN1</i>	no	2 de novo	F
chr2:236,414,455–242,701,103	6,286,648	loss	42	yes	2 de novo	E
chr2:186,674,000–186,771,130	97,130	loss	0	no	1 de novo, 1 inherited	C
chr13:47,048,000–47,569,000	521,000	gain	4	no	2 de novo	F
chr15:18,800,000–29,000,000	10,200,000	gain	42	yes	~1%	A, B, C, E
chr16:29,500,000–30,108,000	608,000	gain	25	yes	~1%	C
chr16:29,500,000–30,108,000	608,000	loss	25	yes	~1%	A, C, E
chr16:60,142,000–61,582,000	1,440,000	loss	<i>CDH8</i>	no	2 de novo	F
chr17:14,040,000–15,410,000	1,370,000	gain	7	yes	1 de novo, 2 inherited	F
chr22:17,080,000–20,200,000	3,120,000	gain	56	yes	4 de novo, 3 inherited	A, C, F

developmental disorders (Kumar et al., 2008; Marshall et al., 2008; Weiss et al., 2008). By examining recurrent de novo CNVs in a large sample of 1,441 multiplex families with ASD, and an additional 299 subjects with autism spectrum disorder, 512 clinical subjects with developmental delay and 18,834 Icelandic controls, Weiss et al. (2008) found a highly significant association of the 16p11.2 microdeletion with ASD. This deletion was detected in approximately 1% of autism and other pervasive developmental disorders, much higher than the frequency observed in the general population which was ~0.01%. Marshall et al. (2008) and Kumar et al. (2008) detected the 16p11.2 deletion at a similar frequency, 2/427 (0.5%) and 4/712 (0.6%) respectively. In all three studies, the reciprocal duplication of 16p11.2 was also observed in ASD at a rate of approximately 1%. This region is now recognized as a ‘hotspot’ (Eichler and Zimmerman, 2008) with classic genomic features of a microdeletion syndrome, such as the presence of tandem segmental duplications at the breakpoints (Stankiewicz and Lupski, 2002; Stankiewicz et al., 2003; Lee and Lupski, 2006). Interestingly, the clinical phenotypes of patients with 16p11.2 microdeletions have not shown any obvious commonality (Kumar et al., 2008; Weiss et al., 2008), suggesting that some of the novel genomic disorders being discovered may have considerable variability in phenotype, being often indistinguishable from ‘garden variety’ autism.

Recurrent de novo rearrangements have been identified at other genomic locations, suggesting that more CNV ‘hot spots’ are yet to be discovered. Here, we assembled a list of recurrent de novo CNVs from recent studies of structural variation in ASD including Jaquemont et al. (2006), Sebat et al. (2007), Szatmari et al. (2007), Christian et al. (2008), Marshall et al. (2008), Mefford et al. (2008), and Weiss et al.

(2008) (Table 1). A majority of de novo mutations that have been observed in at least two unrelated individuals are marked by flanking segmental duplications, consistent with sites that have a high structural mutation rate. The events most frequently reported are the previously identified hotspots at 16p11.2 and 15q11–q13. Other locations where recurrent mutations have been reported in ASD include 1q21.1, 17q12 and 22q11.2. It is important to point out that these three sites have also been reported as recurrent mutations in other neurological disorders (Lee and Lupski, 2006), highlighting the phenotypic variability associated with these rearrangements. For instance, 1q21.1 deletions and duplications have been identified in association with a range of pediatric phenotypes including autism, schizophrenia and congenital anomalies (Mefford et al., 2008). The 17p12 duplication is a well documented ‘causal’ mutation for Charcot-Marie Tooth disease type 1A (Lupski et al., 1991), while 22q11.2 deletions are associated with Velocardiofacial syndrome and schizophrenia (Karayiorgou et al., 1995) and de novo duplications of 22q11.2 have been reported in association with variable phenotypes (Ensenauer et al., 2003).

Given the much higher rate of de novo CNVs in cases compared to controls, many of the mutations identified in these studies are likely to play a causal role in ASD. However, given that relatively few have been reported in more than one family, the evidence implicating most individual loci is not unequivocal. In all studies that have so far examined spontaneous CNVs in non-disease samples, de novo events have been consistently reported in approximately 1% of apparently healthy individuals (Redon et al., 2006; Sebat et al., 2007; Xu et al., 2008). Some of these events represent clinically-neutral mutations, or in cases where only cell line-derived DNAs were used, some mutations could be ar-

tifacts due to cell culture. It is also important to consider that, in a small fraction of patients, de novo changes have been detected at multiple locations in the genome (Sebat et al., 2007; Marshall et al., 2008), raising the likely possibility that not all observed lesions in an individual are clinically relevant. In light of these observations, caution must be used in interpreting the clinical relevance of any 'private' mutation, whether it occurs de novo or not.

For the sake of argument, we should also recognize that the association of de novo structural mutation with disease could imply something other than a causal relationship. One alternative hypothesis to consider is whether de novo CNVs could instead be secondary lesions, i.e. mutations that accumulate as a byproduct of the disease process, as in a disease such as cancer. However, the hypothesis that autism could be a 'fragile genome disorder' is not consistent with certain findings, such as the association of de novo CNVs with ASD primarily in simplex families, the reduced gender bias among patients carrying de novo events, and the preponderance of rearrangements that are deletions. In order to reconcile the hypothesis with these data, we must then assume that the primary insult involved is also genetic in nature, occurs sporadically, with increased penetrance in females, and that the secondary lesions (the observed CNVs) occur through a mechanism that strongly favors deletions. By the law of parsimony, we would consider this hypothesis the less likely of the two.

To summarize, the following key lines of evidence support a causal role for rare SVs in the etiology of ASD. (1) De novo CNVs, approximately 100 kb and larger, collectively occur at a much higher rate in patients than in healthy controls and (2) Some of the most frequent de novo mutations show a significant association with ASD in large cohorts. Additional evidence supporting a causal role includes the reduced ratio of males-to-females in patients with de novo CNVs, (<2:1, as compared to 4:1), consistent with these mutations having increased penetrance.

The findings of CNV studies have revealed that spontaneous structural mutations play a more important role in the etiology of ASD than was known from previous studies based on cytogenetics. Since only a fraction of all spontaneous CNVs in ASD are detected with current microarray platforms, the true rate of such mutations could be significantly higher. Furthermore, rare de novo point mutations are known to play a role in a subset of cases (Barker et al., 1987; Verkerk et al., 1991; Yu et al., 1991; Amir et al., 1999). Therefore, structural rearrangements and point mutations that arise spontaneously could collectively contribute in a large fraction of cases.

Implications for the observed heritability of autism

An important role for de novo mutation has clear implications for genetic models of ASD. The heritability of ASD is estimated primarily based on rates of disease concordance in monozygotic (MZ) and dizygotic (DZ) twins, which are >70% and <10% respectively (Freitag, 2006). The

former reflects both inherited and spontaneous genetic influences, because de novo mutations that occur in the parental germline will invariably be shared between MZ twins. In contrast to earlier genetic models of ASD that accounted for the discrepancy between MZ- and DZ-twin concordance rates by an oligogenic model (Pickles et al., 1995; Constantino and Todd, 2000), new models have been put forward recently that account for spontaneous mutation (Jiang et al., 2004; Zhao et al., 2007). Zhao and Wigler postulated an alternative genetic model that takes de novo mutation into account (Zhao et al., 2007). This model posits that the majority of autism cases occur sporadically as a result of de novo mutation, and it further predicts that, due to variable expression of a high-penetrance allele, some carriers will be mildly affected or not at all. Variable expressivity would enable the transmission of a high-penetrance risk allele to multiple siblings in a subsequent generation. Zhao et al. (2007) examined data on the segregation of autism in multiplex families, specifically focusing on the rates of concordance in children that were born after one pair of concordant siblings was already diagnosed. The authors found that, once two siblings in a family were diagnosed with ASD, the rates of concordance in subsequent children born were ~50% for male siblings and ~20% in female siblings. These data were consistent with the model of alleles having dominant effects in males with reduced penetrance in females.

Although the importance of rare alleles cannot be emphasized enough, genetic models are meant to be a simplification of reality. Rare variant and common variant models are not mutually exclusive, and the true genetic architecture of ASD probably consists of a mixture of modalities including dominant, recessive and low-penetrance alleles. Indeed, rare recessive alleles may explain the increased rate of ASD in consanguineous families (Morrow et al., 2008). In addition common variation could in part explain the variable expressivity observed for some mutations with apparently large effects (Mefford et al., 2008).

Clinical implications of CNV findings

The use of microarray CGH to identify of submicroscopic chromosomal abnormalities in ASD has obvious implications for clinical diagnostics. However, this new technology presents new challenges. As CNV detection platforms become more powerful, the interpretation of CNV data becomes more complex. In order to distinguish between pathogenic mutations and the vast majority of naturally-occurring CNVs that are clinically neutral, it is necessary to demonstrate an association with disease. Size, rarity and de novo occurrence have long been a standard 'criteria for causation' for chromosomal abnormalities detected by cytogenetics. For submicroscopic rearrangements, however, pathogenic mutations may be similar in size to non-pathogenic variants. And while de novo CNVs occur with much higher frequency in ASD than in healthy individuals, de novo occurrence alone is not unequivocal evidence for causation. Furthermore, recurrent CNVs with strong associations with

disease are not perfect predictors of disease outcome, as evident from the variable clinical presentation of some novel genomic disorders (Mefford et al., 2008). Thus, the new cytogenetics must establish new criteria. For instance, rare CNVs including de novo mutations should be implicated with statistical evidence. Determining a prognosis for a carrier should be based on the known penetrance of that mutation for ASD and for other clinical phenotypes associated with the same CNV.

Potential implications for the epidemiology of ASD

The prevalence of autism in the population is matter of great concern to science and to the public. Therefore, we must consider what implications these findings have for the epidemiology of ASD. Two central questions raised by CNV studies are: can structural mutation rates, at individual loci or genome-wide, vary between populations? Can rates of structural mutation fluctuate over time?

The mutation rate at some individual loci can differ between populations due to genetic variation. Rates of de novo mutation are locus specific and are dependent largely on local architectural features of the genome, particularly segmental duplications. Since these architectural features are themselves subject to polymorphism, common structural polymorphisms may contribute to differences in structural mutation rate between common haplotypes. Furthermore, differences in haplotype frequency between populations may lead to population differences in the mutation rate in that genomic region. This is perhaps best exemplified by the recently discovered 17q21.3 microdeletion syndrome (Koolen et al., 2006; Sharp et al., 2006; Shaw-Smith et al., 2006; Varela et al., 2006), where the mutation is associated with an inversion polymorphism that is common in Europeans and rare in other populations (Stefansson et al., 2005). Microdeletion syndromes are almost invariably located in complex regions of the genome, often containing other common inversions and copy number polymorphisms (Mefford et al., 2008), thus, there may exist other recurrent de novo mutations that are frequent in one population and not in others.

Another important epidemiological question is whether genome-wide rates of de novo mutation could be influenced by non-genetic factors such as parental age. Advanced parental age has become established as a risk factor for ASD in offspring (Gillberg, 1980; Reichenberg et al., 2006; Cantor et al., 2007; Croen et al., 2007), with the observed effect strongest for older fathers (Reichenberg et al., 2006). This effect could be explained by genetic or epigenetic factors. Rates of structural and numerical chromosome abnormalities in sperm have been shown to increase with paternal age (Sartorelli et al., 2001; Buwe et al., 2005), raising the interesting hypothesis that spontaneous heritable changes such as de novo CNVs could be a mechanism to explain the paternal age effect. However, strong evidence linking paternal age effects to mutation in the male germline is lacking. Even in disorders that are clearly caused by genetic changes and

which show clear paternal age effect, such as achondroplasia (Tiemann-Boege et al., 2002), the paternal age effect cannot be explained entirely by increased mutation rates in sperm. In order to establish a link between paternal age and the rates of de novo CNVs in autism, both must be examined directly in a large family sample.

Progress toward definitive identification of autism genes

Definitive identification of a CNV risk factor pinpoints a narrow region of the genome that is likely to contain autism genes. The next step is to determine which of the genes in the local region are functionally relevant. This is not a trivial matter, because the frequent recurrent mutations that have so far been identified are relatively large in size (>500 kb) and contain dozens of genes. Additional genetic or functional evidence is needed to confirm the biological relevance of an individual gene. More intensive genetic analysis of candidates in regions identified from large CNV regions may help to pinpoint the biologically relevant factors.

Several of the de novo CNVs identified in recent studies are smaller (≤ 100 kb), and include a few genes or even a single gene. Although the frequencies of these variants appear to be lower than for the previously described 'hot spots' and the corresponding associations are not statistically significant, converging lines of evidence from multiple studies is emerging for some individual genes. For instance, by sequencing and cytogenetic analysis of candidate genes within the 22q13 microdeletion region (Manning et al., 2004), Durand et al. (2007) found additional mutations implicating *SHANK3*. Multiple studies now find evidence for the role of rare neurexin-1 mutations in neuropsychiatric disorders (Kim HG et al., 2008; Kirov et al., 2008; Marshall et al., 2008; Yan et al., 2008). As these examples illustrate, the most biologically informative findings do not necessarily come from the mutations observed most frequently. The rarest alleles, i.e. those that are observed in <0.1% of the sample, may have striking phenotypic effects in individual patients, and may collectively account for a substantial proportion of disease risk in the overall population. This notion is supported by evidence from studies of ASD and schizophrenia, where the majority of de novo mutations reported to date have been observed in single families (Sebat et al., 2007; Xu et al., 2008) and the increased mutational burden of rare structural variants observed in patients with schizophrenia (Stone et al., 2008; Walsh et al., 2008) is greatest for variants observed only once in the sample (Stone et al., 2008). Thus, the rarest of de novo mutations could also serve to identify candidate genes, the most attractive candidates being the genes that are uniquely disrupted by a small de novo deletion or balanced rearrangement. We reviewed the literature and the Autism Chromosome Rearrangement Database to assemble a list of de novo CNVs and de novo balanced cytogenetic rearrangements that disrupt individual genes (Table 2). This list consists of 19 mutations involving 21 differ-

Table 2. Individual genes disrupted by de novo CNVs and balanced rearrangements in ASD. Single gene disruptions caused by de novo deletions, duplications, inversions, and balanced translocations were compiled from the literature and the Autism Chromosome Rearrangement Database. Each entry in the table lists the type of chromosomal aberration whether it is gain, loss, inversion or balanced translocation. It also includes whether the variation was detected by array-CGH or by cytogenetics, the karyotype of the individual the variation was detected in, which study reported the variation, the gene affected, the function of the gene if available, and any other diseases known to be caused by the gene. References are listed by first author (Bakkaloglu et al., 2008; Bonaglia et al., 2006; Borg et al., 2002; Castermans et al., 2003; Ishikawa-Brush et al., 1997; Marshall et al., 2008; Roohi et al., 2008; Sebat et al., 2007; Sultana et al., 2002; Szatmari et al., 2007).

Type of rearrangement	Detection method	Karyotype	Number of cases	Studies	Gene symbol(s)	Candidate gene name	Gene function
loss	aCGH	del(16)(p13.3p13.3)	1	Sebat	<i>A2BP1</i>	Ataxin 2-binding protein 1	Neuronal splicing enhancer
loss	aCGH	del(16)(q24.3q24.3)	1	Marshall	<i>ANKRD11</i>	Ankyrin repeat domain 11	Regulation of transcription
balanced	cytogenetics	t(7;20)(q11.2;p11.2)	1	Sultana	<i>AUTS2</i>	Autism susceptibility gene 2 protein	Unknown
balanced	cytogenetics	inv(7)(q11.22q35)	1	Bakkaloglu	<i>CNTNAP2</i> , <i>AUTS2</i>	Contactin associated protein-like 2	Neuronal cell adhesion molecule
gain	aCGH	dup(8)(p23.2p23.2)	1	Szatmari	<i>CSMD1</i>	Cub and sushi multiple domains 1	May regulate inflammation in the brain
gain	aCGH	dup(8)(p23.3p23.3)	1	Marshall	<i>DLGAP2</i>	Discs large-associated protein 2	Neuronal cell signaling
loss	aCGH	del(7)(q36.2q36.2)	1	Marshall	<i>DPP6</i>	Dipeptidyl-peptidase 6	Modulation of calcium channels
loss	aCGH	del(1)(p21.3p21.3)	1	Marshall	<i>DPYD</i>	Dihydropyrimidine dehydrogenase	Pyrimidine metabolism
gain	aCGH	dup(3)(p14.2p14.2)	1	Sebat	<i>FHIT</i>	Fragile histidine triad gene	Tumor suppressor
loss	aCGH	del(3)(p14.2p14.2)	1	Sebat	<i>FHIT</i>	Fragile histidine triad gene	Tumor suppressor
loss	aCGH	del(7)(p21.1p21.1)	1	Sebat	<i>FLJ16237</i>	Hypothetical protein LOC392636	Putative sterol desaturase
balanced	cytogenetics	t(X;8)(p22.13;q22.1)	1	Ishikawa-Brush	<i>GRPR</i>	Gastrin releasing peptide receptor	Regulates gastrointestinal and central nervous systems function
balanced	cytogenetics	t(5;13)(q12.1;q13.2)	1	Castermans	<i>NBEA</i>	Neurobeachin	Synaptic transmission
loss	cytogenetics	del(2)(p16.3p16.3)	2	Szatmari	<i>NRXN1</i>	Neurexin 1-beta precursor	Neurotransmitter release
loss	cytogenetics	t(2;8)(q36;q21.3)	1	Borg	<i>PAX3</i> , <i>MMP16</i>	Paired box gene 3	Regulation of transcription
balanced	cytogenetics	t(2;9)(p13;p24)	1	Roohi	<i>RAB11FIP5</i>	RAB11 family interacting protein 5 (class I)	Regulates apical membrane trafficking
loss	cytogenetics	t(12;22)(q24.1;q13.3)	1	Bonaglia	<i>SHANK3</i>	Proline-rich synapse-associated protein 2	Synaptic scaffolding protein
loss	aCGH	del(2)(q24.2q24.2)	1	Sebat	<i>SLC4A10</i>	Solute carrier family 4, sodium bicarbonate	pH homeostasis in the central nervous system
loss	aCGH	del(16)(q21q21)	2	Szatmari	<i>CDH8</i>	Cadherin 8, type 2 preproprotein	mediate calcium-dependent cell-cell adhesion

ent genes. These genes include several that play key roles in neurodevelopment, such as *A2BP1* (Nakahata and Kawamoto, 2005) a neuronal splicing enhancer, *DPP6* (Nadal et al., 2003; Kim J et al., 2008) a necessary component of type A potassium channels that regulates neuronal excitability, and *CNTNAP2*, a member of the neurexin superfamily that has been previously implicated in familial epilepsy (Strauss et al., 2006) and Gilles de la Tourette syndrome (Verkerk et al., 2003). Given their established neuronal functions, these genes make excellent candidates for further study. It is likely that some autism genes will harbor additional risk alleles. Indeed, analysis of rare (Bakkaloglu et al., 2008) and common alleles (Alarcon et al., 2008; Arking et al., 2008) of the *CNTNAP2* gene further support its role in ASD. Thus, a more intensive genetic analysis of these and other gene candidates could help to confirm the role of some of these genes in ASD (Abrahams and Geschwind, 2008).

A model for how to use next generation sequencing platforms

CNV-based studies of ASD serve as a model for how emerging technologies for mutation detection, including next-generation sequencing platforms (Mardis, 2008), can be applied to genetic studies of disease. As rapidly evolving technologies bring the field ever closer to the critical ‘\$1000 genome’ milestone, we can realistically envisage applying genome-wide screens for de novo mutation (Sebat et al., 2007; Xu et al., 2008) or mutational burden analysis (Stone et al., 2008; Walsh et al., 2008) to a nearly-complete inventory of SNPs and SVs. Structural variants could represent a substantial fraction of all causal variants in ASD. However, a sequence variant, such as frameshift or stop mutation can be as deleterious to gene function as a large deletion (Durand et al., 2007). Therefore, it is clear that some proportion

of causal variants will be de novo or inherited single nucleotide changes. Thus, further genetic evidence for the role of individual genes, initially identified from CNV studies, can be sought by sequencing candidate genes in additional patient samples. Further studies that examine gene candidates more intensively using next-generations sequencing technologies will be needed to more completely understand the contribution of rare variants and the role of spontaneous mutation in ASD.

Acknowledgement

We wish to thank Michael Wigler, Mary-Claire King, Deborah Levy and James Watson for helpful discussions and support.

References

- Abrahams BS, Geschwind DH: Advances in autism genetics: on the threshold of a new neurobiology. *Nat Rev Genet* 9:341–355 (2008).
- Alarcon M, Cantor RM, Liu J, Gilliam TC, Geschwind DH: Evidence for a language quantitative trait locus on chromosome 7q in multiplex autism families. *Am J Hum Genet* 70:60–71 (2002).
- Alarcon M, Abrahams BS, Stone JL, Duvall JA, Perederiy JV, et al: Linkage, association, and gene-expression analyses identify *CNTNAP2* as an autism-susceptibility gene. *Am J Hum Genet* 82:150–159 (2008).
- Amir RE, Van den Veyver IB, Wan M, Tran CQ, Francke U, Zoghbi HY: Rett syndrome is caused by mutations in X-linked *MECP2*, encoding methyl-CpG-binding protein 2. *Nat Genet* 23:185–188 (1999).
- Arking DE, Cutler DJ, Brune CW, Teslovich TM, West K, et al: A common genetic variant in the neurexin superfamily member *CNTNAP2* increases familial risk of autism. *Am J Hum Genet* 82:160–164 (2008).
- Auranen M, Vanhala R, Varilo T, Ayers K, Kempas E, et al: A genomewide screen for autism-spectrum disorders: evidence for a major susceptibility locus on chromosome 3q25–27. *Am J Hum Genet* 71:777–790 (2002).
- Autism and Developmental Disabilities Monitoring Network Surveillance Year 2002 Principal Investigators; Centers for Disease Control and Prevention: Prevalence of autism spectrum disorders – autism and developmental disabilities monitoring network, 14 sites, United States, 2002. *MMWR Surveill Summ* 56:12–28 (2007).
- Bailey A, Le Couteur A, Gottesman I, Bolton P, Simonoff E, Yuzda E, Rutter M: Autism as a strongly genetic disorder: evidence from a British twin study. *Psychol Med* 25:63–77 (1995).
- Bakkaloglu B, O'Roak BJ, Louvi A, Gupta AR, Abelson JF, et al: Molecular cytogenetic analysis and resequencing of contactin associated protein-like 2 in autism spectrum disorders. *Am J Hum Genet* 82:165–173 (2008).
- Barker D, Wright E, Nguyen K, Cannon L, Fain P, et al: Gene for von Recklinghausen neurofibromatosis is in the pericentromeric region of chromosome 17. *Science* 236:1100–1102 (1987).
- Baron-Cohen S, Belmonte MK: Autism: a window onto the development of the social and the analytic brain. *Annu Rev Neurosci* 28:109–126 (2005).
- Bonaglia MC, Giorda R, Mani E, Aceti G, Anderlid BM, et al: Identification of a recurrent breakpoint within the *SHANK3* gene in the 22q13.3 deletion syndrome. *J Med Genet* 43:822–828 (2006).
- Borg I, Squire M, Menzel C, Stout K, Morgan D, et al: A cryptic deletion of 2q35 including part of the *PAX3* gene detected by breakpoint mapping in a child with autism and a de novo 2;8 translocation. *J Med Genet* 39:391–399 (2002).
- Buwe A, Guttenbach M, Schmid M: Effect of paternal age on the frequency of cytogenetic abnormalities in human spermatozoa. *Cytogenet Genome Res* 111:213–228 (2005).
- Cantor RM, Yoon JL, Furr J, Lajonchere CM: Paternal age and autism are associated in a family-based sample. *Mol Psychiatry* 12:419–421 (2007).
- Castermans D, Wilquet V, Parthoens E, Huysmans C, Steyaert J, et al: The neurobeachin gene is disrupted by a translocation in a patient with idiopathic autism. *J Med Genet* 40:352–356 (2003).
- Christian SL, Brune CW, Sudi J, Kumar RA, Liu S, et al: Novel submicroscopic chromosomal abnormalities detected in autism spectrum disorder. *Biol Psychiatry* 63:1111–1117 (2008).
- Constantino JN, Todd RD: Genetic structure of reciprocal social behavior. *Am J Psychiatry* 157:2043–2045 (2000).
- Croen LA, Najjar DV, Fireman B, Grether JK: Maternal and paternal age and risk of autism spectrum disorders. *Arch Pediatr Adolesc Med* 161:334–340 (2007).
- Durand CM, Betancur C, Boeckers TM, Bockmann J, Chaste P, et al: Mutations in the gene encoding the synaptic scaffolding protein *SHANK3* are associated with autism spectrum disorders. *Nat Genet* 39:25–27 (2007).
- Eichler EE, Zimmerman AW: A hot spot of genetic instability in autism. *N Engl J Med* 358:737–739 (2008).
- Ensenauer RE, Adeyinka A, Flynn HC, Michels VV, Lindor NM, et al: Microduplication 22q11.2, an emerging syndrome: clinical, cytogenetic, and molecular analysis of thirteen patients. *Am J Hum Genet* 73:1027–1040 (2003).
- European Chromosome 16 Tuberous Sclerosis Consortium: Identification and characterization of the tuberous sclerosis gene on chromosome 16. *Cell* 75:1305–1315 (1993).
- Freitag CM: The genetics of autistic disorders and its clinical relevance: a review of the literature. *Mol Psychiatry* 12:2–22 (2006).
- Ghebranious N, Giampietro PF, Westbrook FP, Rezakalla SH: A novel microdeletion at 16p11.2 harbors candidate genes for aortic valve development, seizure disorder, and mild mental retardation. *Am J Med Genet A* 143:1462–1471 (2007).
- Gillberg C: Maternal age and infantile autism. *J Autism Dev Disord* 10:293–297 (1980).
- IMGSAC 2001: A genomewide screen for autism: strong evidence for linkage to chromosomes 2q, 7q, and 16p. *Am J Hum Genet* 69:570–581 (2001).
- Ishikawa-Brush Y, Powell JF, Bolton P, Miller AP, Francis F, et al: Autism and multiple exostoses associated with an X;8 translocation occurring within the *GRPR* gene and 3' to the *SDC2* gene. *Hum Mol Genet* 6:1241–1250 (1997).
- Jacquemont ML, Sanlaville D, Redon R, Raoul O, Cormier-Daire V, et al: Array-based comparative genomic hybridisation identifies high frequency of cryptic chromosomal rearrangements in patients with syndromic autism spectrum disorders. *J Med Genet* 43:843–849 (2006).
- Jamain S, Quach H, Betancur C, Rastam M, Colin-eaux C, et al: Mutations of the X-linked genes encoding neuroligins *NLGN3* and *NLGN4* are associated with autism. *Nat Genet* 34:27–29 (2003).
- Jiang YH, Sahoo T, Michaelis RC, Bercovich D, Bressler J, et al: A mixed epigenetic/genetic model for oligogenic inheritance of autism with a limited role for *UBE3A*. *Am J Med Genet A* 131:1–10 (2004).
- Karayorgou M, Morris MA, Morrow B, Shprintzen RJ, Goldberg R, et al: Schizophrenia susceptibility associated with interstitial deletions of chromosome 22q11. *Proc Natl Acad Sci USA* 92:7612–7616 (1995).
- Kim HG, Kishikawa S, Higgins AW, Seong IS, Donovan DJ, et al: Disruption of neurexin 1 associated with autism spectrum disorder. *Am J Hum Genet* 82:199–207 (2008).
- Kim J, Nadal M, Clemens AM, Baron M, Jung SC, et al: The Kv4 accessory protein *DPPX* is a critical regulator of membrane excitability in hippocampal CA1 pyramidal neurons. *J Neurophysiol* 100:1835–1847 (2008).
- Kirov G, Gumus D, Chen W, Norton N, Georgieva L, et al: Comparative genome hybridization suggests a role for *NRXN1* and *APBA2* in schizophrenia. *Hum Mol Genet* 17:458–465 (2008).
- Koolen DA, Vissers LE, Pfundt R, de Leeuw N, Knight SJ, et al: A new chromosome 17q21.31 microdeletion syndrome associated with a common inversion polymorphism. *Nat Genet* 38:999–1001 (2006).
- Kumar RA, KaraMohamed S, Sudi J, Conrad DF, Brune C, et al: Recurrent 16p11.2 microdeletions in autism. *Hum Mol Genet* 17:628–638 (2008).
- Laumonnier F, Bonnet-Brilhault F, Gomot M, Blanc R, David A, et al: X-linked mental retardation and autism are associated with a mutation in the *NLGN4* gene, a member of the neuroligin family. *Am J Hum Genet* 74:552–557 (2004).

- Lee JA, Lupski JR: Genomic rearrangements and gene copy-number alterations as a cause of nervous system disorders. *Neuron* 52:103–121 (2006).
- Liu J, Nyholt DR, Magnussen P, Parano E, Pavone P, et al: A genomewide screen for autism susceptibility loci. *Am J Hum Genet* 69:327–340 (2001).
- Lohmueller KE, Pearce CL, Pike M, Lander ES, Hirschhorn JN: Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. *Nat Genet* 33:177–182 (2003).
- Lupski JR, de Oca-Luna RM, Slaugenhaupt S, Pentao L, Guzzetta V, et al: DNA duplication associated with Charcot-Marie-Tooth disease type 1A. *Cell* 66:219–232 (1991).
- Manning MA, Cassidy SB, Clericuzio C, Cherry AM, Schwartz S, et al: Terminal 22q deletion syndrome: a newly recognized cause of speech and language disability in the autism spectrum. *Pediatrics* 114:451–457 (2004).
- Mardis ER: The impact of next-generation sequencing technology on genetics. *Trends Genet* 24:133–141 (2008).
- Marshall CR, Noor A, Vincent JB, Lionel AC, Feuk L, et al: Structural variation of chromosomes in autism spectrum disorder. *Am J Hum Genet* 82:477–488 (2008).
- Mefford HC, Sharp AJ, Baker C, Itsara A, Yiang Z, et al: Recurrent rearrangements of chromosome 1q21.1 and variable pediatric phenotypes. *N Engl J Med* 359:1685–1699 (2008).
- Morrow EM, Yoo SY, Flavell SW, Kim TK, Lin Y, et al: Identifying autism loci and genes by tracing recent shared ancestry. *Science* 321:218–223 (2008).
- Nadal MS, Ozaita A, Amarillo Y, Vega-Saenz de Miera E, Ma Y, et al: The CD26-related dipeptidyl aminopeptidase-like protein DPPX is a critical component of neuronal A-type K⁺ channels. *Neuron* 37:449–461 (2003).
- Nakahata S, Kawamoto S: Tissue-dependent isoforms of mammalian Fox-1 homologs are associated with tissue-specific splicing activities. *Nucleic Acids Res* 33:2078–2089 (2005).
- Pickles A, Bolton P, Macdonald H, Bailey A, Le Couteur A, Sim CH, Rutter M: Latent-class analysis of recurrence risks for complex phenotypes with selection and measurement error: a twin and family history study of autism. *Am J Hum Genet* 57:717–726 (1995).
- Pollack JR, Iyer VR: Characterizing the physical genome. *Nat Genet* 32 Suppl:515–521 (2002).
- Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, et al: Global variation in copy number in the human genome. *Nature* 444:444–454 (2006).
- Reichenberg A, Gross R, Weiser M, Bresnahan M, Silverman J, et al: Advancing paternal age and autism. *Arch Gen Psychiatry* 63:1026–1032 (2006).
- Risch N, Spiker D, Lotspeich L, Nouri N, Hinds D, et al: A genomic screen of autism: evidence for a multilocus etiology. *Am J Hum Genet* 65:493–507 (1999).
- Roohi J, Tegay DH, Pomeroy JC, Burkett S, Stone G, et al: A de novo apparently balanced translocation [46,XY,t(2;9)(p13;p24)] interrupting *RAB11FIP5* identifies a potential candidate gene for autism spectrum disorder. *Am J Med Genet B Neuropsychiatr Genet* 147B:411–417 (2008).
- Sartorelli EM, Mazzucatto LF, de Pina-Neto JM: Effect of paternal age on human sperm chromosomes. *Fertil Steril* 76:1119–1123 (2001).
- Sebat J, Lakshmi B, Malhotra D, Troge J, Lese-Martin C, et al: Strong association of de novo copy number mutations with autism. *Science* 316:445–449 (2007).
- Sharp AJ, Hansen S, Selzer RR, Cheng Z, Regan R, et al: Discovery of previously unidentified genomic disorders from the duplication architecture of the human genome. *Nat Genet* 38:1038–1042 (2006).
- Shaw-Smith C, Redon R, Rickman L, Rio M, Willatt L, et al: Microarray based comparative genomic hybridisation (array-CGH) detects submicroscopic chromosomal deletions and duplications in patients with learning disability/mental retardation and dysmorphic features. *J Med Genet* 41:241–248 (2004).
- Shaw-Smith C, Pittman AM, Willatt L, Martin H, Rickman L, et al: Microdeletion encompassing *MAPT* at chromosome 17q21.3 is associated with developmental delay and learning disability. *Nat Genet* 38:1032–1037 (2006).
- Stankiewicz P, Lupski JR: Genome architecture, rearrangements and genomic disorders. *Trends Genet* 18:74–82 (2002).
- Stankiewicz P, Shaw CJ, Dapper JD, Wakui K, Shaffer LG, et al: Genome architecture catalyzes nonrecurrent chromosomal rearrangements. *Am J Hum Genet* 72:1101–1116 (2003).
- Stefansson H, Helgason A, Thorleifsson G, Steinthorsdottir V, Masson G, et al: A common inversion under selection in Europeans. *Nat Genet* 37:129–137 (2005).
- Stone JL, O'Donovan MC, Gurling H, Kirov GK, Blackwood DH, et al: Rare chromosomal deletions and duplications increase risk of schizophrenia. *Nature* 455:237–241 (2008).
- Strauss KA, Puffenberger EG, Huentelman MJ, Gottlieb S, Dobrin SE, et al: Recessive symptomatic focal epilepsy and mutant contactin-associated protein-like 2. *N Engl J Med* 354:1370–1377 (2006).
- Sultana R, Yu CE, Yu J, Munson J, Chen D, et al: Identification of a novel gene on chromosome 7q11.2 interrupted by a translocation breakpoint in a pair of autistic twins. *Genomics* 80:129–134 (2002).
- Szatmari P, Paterson AD, Zwaigenbaum L, Roberts W, Brian J, et al: Mapping autism risk loci using genetic linkage and chromosomal rearrangements. *Nat Genet* 39:319–328 (2007).
- Tiemann-Boege I, Navidi W, Grewal R, Cohn D, Eskenazi B, Wyrobek A J, Arnheim N: The observed human sperm mutation frequency cannot explain the achondroplasia paternal age effect. *Proc Natl Acad Sci USA* 99:14952–14957 (2002).
- Varela MC, Krepischi-Santos AC, Paz JA, Knijnenburg J, Szuhai K, et al: A 17q21.31 microdeletion encompassing the *MAPT* gene in a mentally impaired patient. *Cytogenet Genome Res* 114:89–92 (2006).
- Verkerk AJ, Pieretti M, Sutcliffe JS, Fu YH, Kuhl DP, et al: Identification of a gene (*FMR-1*) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. *Cell* 65:905–914 (1991).
- Verkerk AJ, Mathews CA, Joosse M, Eussen BH, Heutink P, Oostra BA: *CNTNAP2* is disrupted in a family with Gilles de la Tourette syndrome and obsessive compulsive disorder. *Genomics* 82:1–9 (2003).
- Vorstman JA, Staal WG, van Daalen E, van Engeland H, Hochstenbach PF, Franke L: Identification of novel autism candidate regions through analysis of reported cytogenetic abnormalities associated with autism. *Mol Psychiatry* 11:18–28 (2006).
- Walsh T, McClellan JM, McCarthy SE, Addington AM, Pierce SB, et al: Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. *Science* 320:539–543 (2008).
- Weiss LA, Shen Y, Korn JM, Arking DE, Miller DT, et al: Association between microdeletion and microduplication at 16p11.2 and autism. *N Engl J Med* 358:667–675 (2008).
- Xu B, Roos JL, Levy S, van Rensburg EJ, Gogos JA, Karayiorgou M: Strong association of de novo copy number mutations with sporadic schizophrenia. *Nat Genet* 40:880–885 (2008).
- Yan J, Noltner K, Feng J, Li W, Schroer R, et al: Neurexin 1alpha structural variants associated with autism. *Neurosci Lett* 438:368–370 (2008).
- Yu S, Pritchard M, Kremer E, Lynch M, Nancarrow J, et al: Fragile X genotype characterized by an unstable region of DNA. *Science* 252:1179–1181 (1991).
- Zhao X, Leotta A, Kustanovich V, Lajonchere C, Geschwind DH, et al: A unified genetic theory for sporadic and inherited autism. *Proc Natl Acad Sci USA* 104:12831–12836 (2007).