doi:10.1111/pcn.12128

Psychiatry and Clinical Neurosciences

PCN Frontier Review

Copy-number variation in the pathogenesis of autism spectrum disorder

Emiko Shishido, PhD,[1,2,3] Branko Aleksic, MD, PhD[3] and Norio Ozaki, MD, PhD[3]*

[1]*National Institute for Physiological Sciences,* [2]*Restart Postdoctoral Fellowship of Japan Society for the Promotion of Science, Tokyo, and* [3]*Department of Psychiatry, Nagoya University Graduate School of Medicine, Nagoya, Japan*

Autism spectrum disorder is a neurodevelopmental disorder present in 1% of the population, characterized by impairments in reciprocal social interaction, communication deficits and restricted patterns of behavior. Approximately 10% of the autism spectrum disorder population is thought to have large chromosomal rearrangements. Copy-number variations (CNV) alter the genome structure either by duplication or deletion of a chromosomal region. The association between CNV and autism susceptibility has become more apparent through the use of methods based on comparative genomic hybridization in screening CNV. The nature of the high CNV rate in the human genome is partly explained by non-allelic homologous recombination between flanking repeated sequences derived from multiple copies of transposons or mobile genetic elements. There are hotspots for CNV in the human genome, such as 16p11.2 and 22q11.2. Genes involved in CNV are supposed to have copy-number dose-dependent effects on the behavior of affected individuals. Animal models give insight into the possible interactions between core genetic loci and additional factors contributing to the phenotypes of each individual. If affected genes code for cellular signaling molecules, reducing the dosage in the intracellular signaling pathway may result in the malfunction of the nervous system. The genetic background of autism spectrum disorder is highly heterogenic and most common or rare CNV do not lead to autism spectrum disorders in the majority of cases, but may occasionally increase the risk of developing an autism spectrum disorder.

Key words: autism spectrum disorders, copy-number variation, *de novo* mutation, schizophrenia, signal transduction.

AUTISM IS A neurodevelopmental disorder characterized by impairments in reciprocal social interaction, communication deficits and repetitive and restricted patterns of behavior and interests.[1–3] The three disorders, autism, Asperger's syndrome and pervasive developmental disorder – not otherwise specified (PDD-NOS) differ with regard to symptom severity and early development of language, cognitive and social behavior.[1,4] In the DSM-5, the core symptoms of autism spectrum disorder (ASD) are defined using two categories of behavior: (i) deficit in social communication; and (ii) stereotyped behavior. However, there was significant diversity among the patients diagnosed in terms of severity, intelligence,

*Correspondence: Norio Ozaki, MD, PhD, Department of Psychiatry, Nagoya University Graduate School of Medicine, 65 Tsurumai-cho, Showa-ku, Nagoya, Aichi 466-8550, Japan.
Email: ozaki-n@med.nagoya-u.ac.jp
Accepted 10 October 2013.

Table 1. Recent advances in genetic study of ASD

Year/period	Methods/resources	Events	References
1970–	Twin studies	Strong genetic contribution (60–90% in twin studies)	Hallmayer *et al.*[9]
1998–	Linkage analysis	Polygenic mode of transmission of autism	Freitag[4]
2003	DNA sequencer	Contiguous sequence of human genomic DNA	International Human Genome Sequencing Consortium[10]
2007–	Array CGH	Analysis of *de novo* CNV in autism genome	Sebat *et al.*[11]
2008	1000 genome project	Structural variation of human genomes	Kidd *et al.*[12]
2010	Autism genome project	Hundreds of genes associated with ASD	Anney *et al.*[13]
2013	Various resources	Common variants are the major part of autism genetics	Stein *et al.*[14]

ASD, autism spectrum disorder; CGH, comparative genomic hybridization; CNV, copy-number variation.

and language impairment. In addition, some are characterized by hyperactivity, hyper- or hypo-reactivity to sensory input as well as other behaviors. The individuals show impairments of affected area and atypical development before age 3. The child is occasionally afflicted with temper tantrums, self-destructive behavior and epilepsy. Some children with ASD show delays in motor coordination, which may be apparent from late infancy.[5] In other cases, such physical conditions may not be apparent.

Using the former criteria of the DSM (DSM-IV), patients with ASD could be diagnosed with several different disorders, including autism, Asperger's syndrome and PDD-NOS.[1] However, these categorical diagnoses did not reflect the biological bases of ASD and the boundary between each diagnostic nomenclature became less meaningful even almost 20 years after its publication. Using the new criteria of the DSM-5, ASD is classified under neurodevelopmental disorders and additional characteristics other than the two core symptoms will be added with individualized specifications. For example, genetic conditions, if known, will be added to the patient's diagnosis as a genetic association. Thus, the most recent version of the DSM represents a more accurate description of the spectrum based on both clinical and scientific observation.

Autism was first outlined in 1943 by Leo Kanner, an American professor of child psychiatry educated in Germany. Kanner named the syndrome 'infantile autism'.[4] At about the same time, Hans Asperger, a pediatrician in Austria, identified a similar condition, which is now known as Asperger's syndrome.[6] In the first edition of the DSM, published in 1952, autism was not given its own diagnostic criteria. Instead, children demonstrating autistic-like symptoms were diagnosed with 'childhood schizophrenia'. Schizophrenia and autism remained linked in the minds of many researchers until the 1960s. It was only then that medical professionals began to have a separate understanding of autism in children.[7] In the last 70 years, autism and ASD have gone from being an obscure condition to a familiar diagnosis. The prevalence of autism has been estimated to be 15–20 per 10 000 people, whereas a broader phenotype is seen in approximately 1% of the population.[2] In a 2011 study from South Korea, the observed frequency was about 1 in 40 children, but many of them were not diagnosed nor treated.[8]

For more than half a century, people were trying to discover the cause of autism spectrum disorders, but many of the exact pathoetiological mechanisms behind the symptoms of autism remained elusive. The latest edition of the DSM[2] includes some significant changes to the diagnostic criteria for autism, grouping several previously separate disorders under one umbrella. Previously in the DSM-IV,[1] there were five disorders, each of which had a unique diagnosis: Autistic Disorder or classic autism, Asperger's Disorder, PDD-NOS, Rett's Syndrome, and Childhood Disintegrative Disorder. Although the definition of autism is changing, the core characteristics of the disorder remain the same. As people with all levels of autism display many of the same characteristics, but vary in the degree to which they display them, the new DSM-5 criteria may better reflect that autism is a spectrum, rather than a group of separate disorders.

Table 1 presents early findings from twin studies that have indicated a genetic contribution in ASD.[9–14] Twin studies show a concordance of 60–92% in

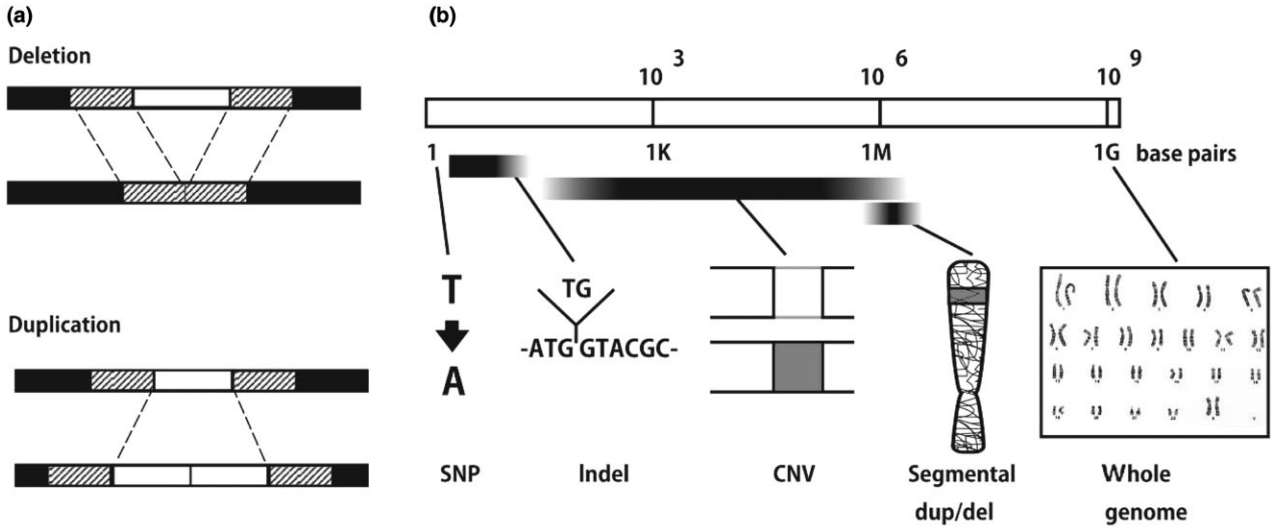

Figure 1. (a) Schematic drawing of a deletion and duplication compared to a reference genome. (b) Classes of variation were shown with size in logarithm. A single nucleotide change is referred to as an single-nucleotide polymorphisms (SNP). Small insertional sequences or deletions are called indels. Indels usually mean a loss or a gain of rather a small size of sequences of up to 100 bp. Copy-number variations (CNV) include duplication or deletions of medium size. CNV of several mega bp are usually called segmental duplications as seen *in situ* as chromosome aberrations.

monozygotic twins and 0–10% in dizygotic pairs.[15,16] Recent studies using the Autism Diagnostic Observation Schedule/Autism Diagnostic Interview, Revised reported that the concordance rate for autism in monozygotic twins was 60–90% and 5–30% in dizygotic twins.[9] Because twins share both environmental and genetic factors, researchers may want to investigate which factor has more influence on the risk for autism. Unique environmental factors are not always measurable, whereas genetic influences can be assessed partly by screening for the genetic variation.

A significant increase in the genome research for ASD in recent years is presented in Table 1. The first comprehensively annotated full-length human genomic sequence was published in 2003.[10] As a result, researchers have become aware of large structural variations within the genomes of each individual.[12,17] Copy-number variations (CNV) are the most common structural variations found in the human genome[16] and represent a major part of genetic variability.[18] As in Figure 1a, CNV include duplications and deletions of the genomic DNA sequence. Compared to the single-nucleotide polymorphisms (SNP) and small indels, CNV affect large portions of the DNA and change the information coded for the relevant genomic region (Fig. 1b). CNV are a remarkably frequent variation, involving more than 10% of the genome.[16] Sebat *et al*. demonstrated that submicroscopic (<500 bp) variations of CNV are widespread in the normal human genome.[17,19] CNV are one of the major sources of person-to-person differences, and on average, there are 1000 CNV in the genome, accounting for around 4 million bp of the genomic difference,[19] although this information may not be accurate due to a detection threshold for the available CNV screening methodologies. The most sensitive commercial oligonucleotide arrays cannot detect CNV smaller than 500 bp.[20]

As research studies focused on elucidating the impact of CNV in order to advance human health, the association between CNV and autism susceptibility has become more apparent (Tables 1 and 2).[9–14,19,21–24] Specifically, in 2007, Szatmari *et al*.[15] searched for CNV in familial ASD and found that inherited CNV may increase the susceptibility of ASD. They also reported CNV that were not transmitted from parents (*de novo* CNV). Furthermore, in 2007, Sebat *et al*. found that *de novo* CNV were associated with ASD.[11,25] More importantly, the association was at a statistically significant level ($P = 0.0005$), and this was one of the first pieces of evidence to show an association between CNV and ASD. Sebat and colleagues compared CNV between patients and unaffected control subjects using a comparative genomic hybridization (CGH)

Table 2. Major CNV with strong association with ASD and developmental disorders

CNV locus	Frequency in ASD[21–23] ($n = 2120$)	Frequency in developmental disorders[24] ($n = 15767$)	Frequency in control[21–24] ($n = 8329$)	Association to ASD	Related disorders[19,21–23]
1q21	0.2%	0.5%	<0.1%	dup	Schizophrenia (deletion)
5p15.2	0.1%	–	0%	del	Novel region found for ASD
7q11.23	0.2%	0.4%	0%	dup	Williams–Beuren syndrome (deletion)
15q11-13	0.1%	0.3%	0%	del, dup	Prader–Willi syndrome (paternal deletion), Angelman syndrome (maternal deletion), ASD (maternal duplication)
15q13	0.2%	0.4%	<0.1%	dup	Schizophrenia (deletion), bipolar disorder (duplication)
16p11.2	0.8%	0.5%	<0.1%	del, dup	Schizophrenia (deletion), bipolar disorder, OCD (duplication)
17p11.2	0.2%	0.2%	0%	dup	Smith–Magenis syndrome (deletion)
22q11.2	0.1%	0.9%	0%	del, dup	DiGeorge syndrome, schizophrenia (deletion)

ASD, autism spectrum disorder; CNV, copy-number variation; OCD, obsessive–compulsive disorder.

array and validated the result using several methods, including tests for parental analysis, cytogenesis and *in situ* hybridization. As many CNV are genetically passed on from parents to children, Sebat and colleagues' strategy allowed them efficient detection of *de novo* CNV. Subsequent studies identified novel CNV associated with ASD.[19,21–23] The major *de novo* CNV associated with ASD are shown in Table 2. The CNV listed in Table 2 are the most frequent, and known as syndromic CNV. Syndromic CNV other than the ones listed above are listed on the DECIPHER website.[26] The regions of CNV listed in Table 2 disrupt a region of DNA ranging from 50 kb up to 5 Mbp. In most cases, the CNV disrupted region contains 10–30 genes. Some of the genes involved in the CNV are supposed to have copy-number dose-dependent effects on the behavior of affected individuals.[27,28] In all the cases, aberrations in the same locus were associated with other psychiatric disorders (Table 2).

CNV ARE FREQUENT AND LARGE CHANGES IN THE HUMAN GENOME

Typically, the human genome contains two copies of genes in autosomal chromosomes. With CNV, the number can vary between zero and four depending on the combination of deletion and duplication.[16] In addition to alterations in the gene copy number, chromosomal rearrangements sometimes cause gene disruption, exon-shuffling or altered gene expressions.[29] The precise number of CNV in the human genome is yet to be discovered, but rough estimation with available methods to date shows that any two genomes may differ by more than 1000 CNV. This means that any two individuals differ in 1% of their chromosomes by either gains or losses in parts of their DNA sequence.

Lupski *et al.* estimated the locus-specific mutation rates of CNV.[30] The mutation rate of CNV is 100 to 1000 times higher than the mutation rates for SNP and different among loci ranging from 1.7×10^{6} to 1.0×10^{4} per locus per generation. CNV can affect a region of several hundred to a few mega bp.[31] The most frequent CNV are less than 1 kb in length.[31] Only a fraction (5–10%) of people have CNV larger than 500 kbp in length.[32] For example, patients with DiGeorge syndrome typically have 1.5–3 Mbp deletion in 22q11.2 and the estimated mutation rate for this locus is 1.25×10^{-4} according to the birth prevalence.

METHODS FOR DETECTING COPY-NUMBER VARIATIONS

Traditionally, fluorescence *in situ* hybridization (FISH) and array comparative genomic hybridization (aCGH) or SNP arrays are used to detect CNV. However, because of their low resolutions (approxi-

Table 3. Population of 16p11.2 CNV in three different reports[†]

	First screening	Cohort size of ASD (control)	16p11.2 CNV in case (del, dup)	16p11.2 CNV in control (del, dup)	Population
Sebat *et al.* 2007[11]1818	ROMA	195 (196)	1 (1del,0dup)	0	–
Pinto *et al.* 2010[22]	Solid phase hybridization	998 (1287)	9 (6del,3dup)	0	European
Sanders *et al.* 2011[21]	Solid phase hybridization	1124 (872)	16 (8del,8dup)	3 (3 dup)	Caucasian, Hispanic, Asian, African American

[†]Only CNV larger than 500 000 base pairs are counted. Both *de novo* and inherited CNV are included. The individual CNV data from SFARI gene.[37]
ASD, autism spectrum disorder; CNV, copy-number variation; ROMA, representational oligonucleotide microarray analysis.

mately 5–10 Mbp for FISH, and 10–25 kbp with 1 million probes for aCGH/SNP array[33]), short CNV are still difficult in terms of detection. Recently, the next-generation sequencing technology advanced breakthroughs and was used in various fields,[34] for example, for the detection of CNV with high resolution (<10 kbp); however, due to the relatively young stage of the methodology, the performance is not yet fully understood.[35]

Methods based on comparative genomic hybridization are commonly used in CNV screening studies.[36] aCGH is a modified hybridization technique using an array DNA probe.[17,21,22] This allows genome-wide detection of CNV of more than a few hundred nucleotides. However, as the most sensitive commercial oligonucleotide arrays cannot detect CNV smaller than 500 bp,[20] the number of CNV per genome might be underestimated due to the inherent limitations of aCGH technology. In other words, each CNV screening study has different thresholds to detect CNV. For example, three reported studies had different numbers of 16p11.2 cases and controls (Table 3).[11,21,22,37] Because of this technical limit of using aCGH, a significant portion of CNV are denoted as false-negatives. For smaller deletions or duplications less than 100 bp, reading depth method using a next-generation sequencer may be applicable.[38] As discussed by Alkan and colleagues, sequencing of the human genome with highly repetitive sequences has limitations in its accuracy.[35] Each research paper utilizes different screening methods to detect CNV, as well as different validation methods. The detection threshold has been improving and recent reports of large-scale studies on CNV[21,22] have achieved a more refined mapping of CNV.

Currently, nearly 10% of the ASD population is thought to have large chromosomal rearrangements, while others may not. Most of the CNV are inherited from parents and others occur *de novo* in germ cells.[16] The formation of the CNV is mediated by several different mechanisms, such as non-allelic homologous recombination (NAHR), nonhomologous end-joining, break-point induced replication and transposon-mediated events. Of these, NAHR accounts for most of the CNV (Fig. 2a).[12,29] As noted above, the largest CNV that has a strong association with ASD is usually found as a recurrent *de novo* CNV. Some, but not all, genomic CNV are generated *de novo* in the germ line. Using trio analysis (Fig. 2b), CNV found in probands, but not in parents, were associated with an increased risk for ASD. In these comparative analyses,[11,22] certain CNV were found only in probands. *De novo* CNV were not found in the somatic cells of parents; instead they were newly formed in the parents' reproductive cells. In a classical

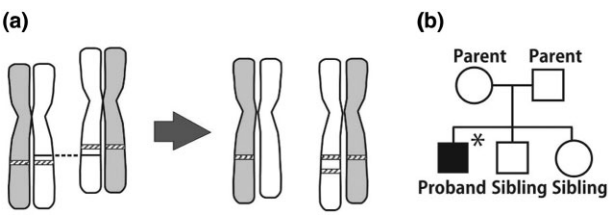

Figure 2. (a) Non-allelic homologous recombination between two chromosomes. The recombination between two homologous regions near the gene (dashed box) occasionally results in the loss or duplication of the gene. (b) Trio analysis to detect *de novo* copy-number variations (CNV). Comparing the CNV of family members.

view of meiosis, chromosomes align precisely at the recombination site and crossovers do not change the number of genes in the recombinant chromosomes. The presence of duplicated sequences on a chromosome increases the frequency of chromosome misalignment during meiosis I. Misalignment can result in unequal crossover events, which increase the copy number of the genes on one recombinant chromosome, while decreasing the copy number of the genes on the other recombinant chromosome (Fig. 2a). Therefore, the misalignment is sometimes associated with the presence of segmental duplications.[39] The rearrangement of chromosomes usually occur between the allelic region of two homologous chromosomes, but recent findings from human genome analyses have shown that the human chromosome frequently undergoes non-allelic recombinations between two chromosomes.[10] Details on this have been described in several reports.[29,40] Some of the hotspots, such as 17p11.2, were investigated in detail.[40] That particular region of chromosome 17 is enriched with repeated sequences, and duplication of this region through NAHR has been associated with ASD. The nature of the high CNV rate in the human genome is partly explained by flanking repeated sequences. In fact, about half of our genomic contents are occupied by repeated sequences. Most of the repeated sequences are multiple copies of transposons or mobile genetic elements, such as L1 and Alu.[41]

BIOLOGICAL MEANINGS OF THE COPY-NUMBER VARIATIONS

Large-scale genome analysis points to and clarifies an association between many neuronal genes in ASD.[20,42,43] Ebert *et al.* explained the reason why neural activity genes are important.[20] Traditional physiology has indicated that neurons change their activity in an excitation-dependent manner. Long-term potentiation and depression accounts for memory and cerebellum learning processes. Synaptic membrane proteins and components of postsynaptic density (PSD) structures have been identified by both genome-wide association and CNV analyses.[20,44] Neurexins are presynaptic membrane proteins, while neuroligins are post-synaptic proteins that function as neurexin receptors.[45] These two types of proteins work together to modulate the formation and function of synapses. Mouse model studies have indicated that a condition without particular types of neurexins and neuroligins may disrupt a specific connection in the brain.[20] The CNV in either of the neurexin or neuroligin loci are rarely (<0.1%) found in the ASD population, which is indicative of the heterogenic nature of the pathogenesis.[20,44] *SHANK* proteins, *SHANK1* to *SHANK3*, function as scaffolds in the PSD of excitatory synapses. The knockout mice of several *SHANK* family proteins exhibit behaviors similar to those observed in ASD, including deficits in social interactions and repetitive behaviors, such as excessive grooming. The *SHANK* proteins, together with other scaffold proteins and certain types of kinases, regulate the organization of postsynaptic structures in an activity-dependent manner.

Common proteins are also encoded by ASD-related genes, such as Na+/H+ exchanger, and cell adhesion molecules, such as contactin, protocadherin 10 and contactin-associated protein-like 2.[42] Genes concerning neuronal activity regulations have also been implicated. Those include Fragile mental retardation 1, Methyl CpG binding protein, deleted in autism 1, Ataxin 2-binding protein 1 and ubiquitin protein ligase (*UBE3A*). *UBE3A* is an enzyme involved in protein degradation. It attaches to ubiquitin to any proteins that should be degraded.[44] Angelman syndrome is characterized by motor dysfunction, speech impairment, seizures and a high prevalence of autism. The syndrome is caused by deletions of the maternally inherited chromosomal region of 15q11-q13. *UBE3A* is included in the deleted region and in one of the candidate genes that cause the phenotypic feature of the 15q11-q13 CNV. One of the roles of the ubiquitin pathway mediated by *UBE3A* is degrading a portion of structural protein in neuronal cells and regulating the turnover of synaptic components. In mice carrying a deletion of *UBE3A* gene (maternal copy), maturation of excitatory neurons in the visual cortex was impaired and the postnatal development of the neuronal circuit was defective.[46] In the same mice, inhibitory input into neocortical pyramidal neurons was diminished owing to the reduced vesicle cycling of the interneuron in the brain region.[47] These studies suggest that the balance between excitatory and inhibitory neurons is disturbed when removing a maternal copy of the *UBE3A* gene.

HOW ABOUT SMALLER CNV?

Girirajan *et al.* examined the global tendency of CNV size in an ASD population and reported the fact that children with ASD have an elevated frequency of

CNV.[48] There are hundreds of genes implicated in the cause of ASD;[42] however, many smaller CNV, which may disrupt or duplicate one or two genes, have not been found to be statistically significant in previous major reports. CNV that affect genes previously implicated in ASD have low penetrance (e.g. *SHANK3*), suggesting that most of the smaller CNV may not be pathological. However, each report[21,22] consists of about 2000 patients and a comparable number of control groups. Furthermore, the event number (CNV found in a population) for each genetic loci is usually very small, in many cases as small as just one or two. The rarity of the event in one report makes it difficult to evaluate each small CNV locus and to estimate its contribution to ASD. In a report investigating the genomic aberrations related to schizophrenia, Gilman *et al*. used cluster analysis to analyze how genetic loci are interconnected in terms of their biological function.[49] This type of analysis integrates many sources of evidence for protein interactions (proteins are selected because they are disrupted by small CNV), using them as features for a classifier of interactions/non-interactions.[22] These computational methods help us understand the nature of each contributing genetic loci in the pathology of ASD. However, researchers still need a large (worldwide) collection of data in order to achieve statistical significance.

The reason why those genes are accompanied by ASD is that ASD is not caused by a single gene mutation. This is driven by an analogy of the fragile-X syndrome,[16,20] which develops in a person who has the fragile-X chromosome with a large number of triplet repeats, which change the DNA's structure. This chromosome is found in approximately 3000–7000 people and, among them, only one-quarter of the male subjects carrying fragile-X develop ASD. The rest of the male subjects with the fragile-X chromosome do not meet the criteria for ASD. This is explained by the fact that while fragile-X is a major factor, other genetic loci are also important contributors in the development of ASD. Similarly, this may be the case for the major CNV previously mentioned. Stein *et al*. documented the current understanding of genome analysis on ASD[14] and summarized the total contribution of CNV, combining the results from several large-scale cohort analyses. According to their analysis, the largest genetic contributor for the development of ASD was common CNV and SNP. The common variations (found in >1–5% of the total population) are the largest driving factor for ASD.[14] However, each common variation is supposed to have only a minor effect, but a number of them together in one person may exhibit a larger effect. In contrast, among the ASD risk factors within a population, *de novo* CNV only have minor effects on ASD.

MODELS OF OLIGOGENIC INTERACTION

There are as many as 200 genes implicated in the pathogenesis of ASD. This covers approximately 1% of the functional genes that humans have. Most of this evidence is coming from the association analysis of SNP found in ASD patients. However, because the genetic background of ASD is highly heterogenic, the mutation frequency for each genetic element or gene is very low in ASD. Many of the implicated genes code for cellular signaling molecules, and reducing the dosage in the pathway may possibly result in malfunction of the nervous system. Usually the signaling pathways are multivalent and not just a case of a simple model. Therefore, this makes it difficult for future research to elucidate the genetic cause of ASD. One approach may be to start with a selected mutation to search for interactions between other genetic loci. A good place to start is the 16p11.2 CNV.[50] The 16p11.2 is the most frequent CNV and the population of ASD carrying 16p11.2 CNV is nearly 1 in 100. The 16p11.2 CNV is also found in approximately 1 in 1000 within a non-ASD control population. One of the more interesting issues with 16p11.2 is that it is also related to other psychiatric conditions, such as schizophrenia, bipolar disorder, obsessive–compulsive disorder (OCD) and obesity. The 16p11.2 region contains 29 genes and is 600 kb in length.[51] Thus, investigating an individual's trait and genetic background will help to understand a second genetic locus, which may affect the phenotype. Two hundred families in North America have already joined the cohort study.[27] In addition, there is a mouse model for 16p11.2. Half of the 16p11.2 deletion mice die postnatally. Mice that survive to adulthood are healthy and fertile, but have alterations in their hypothalamus and exhibit a 'behavior trap' phenotype.[27] These animal models may also give insight into the possible interaction between the core genetic loci and additional factors contributing to the phenotypes of each individual.

The 22q11.2 CNV is known as the 22q11.2 syndrome or DiGeorge syndrome for the deletion phenotype. The syndrome is known as a genetic disorder caused by hemizygous deletions on the locus, with a

population prevalence of 1 in 4000.[52] The deleted syndrome is associated with an elevated rate of schizophrenia and other psychiatric conditions. In addition, it has certain characteristic features, including heart defects, facial anomalies, and T-cell immunodeficiency. The physical and behavioral (i.e. psychiatric) phenotype prominent in the 22q11.2 syndrome is seen as a variable expression and combination, each with different penetrance. The 22q11.2 locus contains a prominent repeated region flanking the CNV region. Both the deletion and duplications are generated by the NAHR between the repeats. The duplication of this locus is associated with ASD, but the penetrance is low.[21,24,37] The 22q11.2 deletion is usually found as *de novo* CNV (>90%) and is associated in almost every individual with any psychotic phenotype.[53] The phenotypes associated with 22q11.2 deletion are highly variable; 20–25% is associated with schizophrenia,[53] mood disorders,[54] ADHD,[54] OCD[54] and learning difficulties.[53] Children with 22q11.2 deletion syndrome are found with ASD with varied frequency.[54–56] The sub-threshold autistic symptoms may be more common in children with the 22q11.2 deletion syndrome. Taken together, the biological nature of ASD and schizophrenia in patients with 22q11.2 CNV may share the same genetic component.

The genetic evidence for CNV in the ASD population is a strong indication of influences of gene dosage in the pathogenesis. But each group of people with the same CNV (as listed in Table 2) shows a variety of phenotypes in terms of psychosis, intellect, or additional physical features, such as obesity and heart disease. Studies on a group who carried a deletion in the same locus indicated that other mutations additively influenced their intellectual and disease conditions.[57] More generally, the genetic interaction between mutations is reported in experimental animals. One example is genes coding for the Notch family protein.[58,59] The Notch family protein was originally identified in insects, and its mutation shows a variety of interactions between several genetic loci in a dose-sensitive manner. The other is an experiment targeted to find genetic loci interacting with a cell-surface receptor.[60,61] In this sensitive genetic condition, they searched for interacting loci through a change in photoreceptor cells of an insect's eye. Five such loci are identified and one of them is encoding Ras, an important signaling molecule for cytoskeletal reorganization in many cell processes (Fig. 3a). In certain genetic conditions, reducing the number of interacting genes influenced the penetration of phenotype. Additionally, these influences are

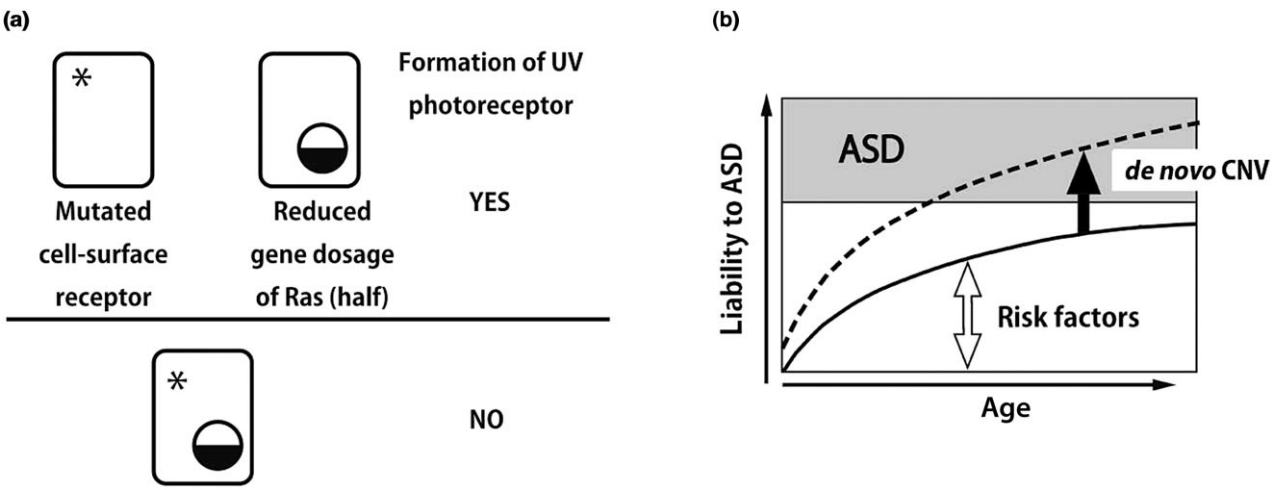

Figure 3. (a) Schematic illustration to show the genetic interaction in a dose-sensitive manner. In a study of insects where genetic approach is available, gene dosage of Ras can affect the loss of UV photoreceptor cells. This is seen in a condition where gene coding for cell-surface receptor kinase is mutated by molecular engineering. This example indicates the specific gene–gene interaction in dose-sensitive manner for loss of a specific cell-type. (b) A model to explain the effect of *de novo* copy-number variations (CNV) in the pathogenesis of autism spectrum disorder (ASD). Each individual has inherited genetic background with varied potential for developing a specific behavioral trait. A large *de novo* CNV may raise the possibility of ASD. For those who are diagnosed later in childhood or in adulthood, the diagnosis may differ from ASD. For details, see the section in the Discussion on 22q11.2 CNV and ASD.

only observed in a specific combination of a mutant and a locus in a dose-dependent manner. That means that the interaction between secondary genetic loci can cause different effects on individuals. In such a condition, the phenotype is seen only in a certain combination of genes in a dose-sensitive manner. The majority of common and rare genetic variants that one person has in most of the cases do not lead to ASD, but in certain cases lead to increased risk of ASD as well as a high risk of other disorders (Fig. 3b).

CONCLUSIONS

Judging by the studies on large CNV above, there seems to be a population of people who carry the same CNV, but do not exhibit any definitive psychiatric features. The genes in CNV are responsible for biological expressions, such as protein dosage or mRNA expression levels. However, due to the limits of interrogating resolution and genome coverage of the present technologies used in CNV studies, many more undiscovered CNV may exist in the human genome, and more comprehensive studies are expected to advance our knowledge of the distribution, formation, and genetic susceptibility of CNV. Moreover, psychiatric disorders are only defined by the pattern of behavior. Table 2 presents individuals with a certain CNV (e.g. 16p11.2 and 22q11.2) who have an increased probability of ASD. Some large CNV may raise the possibility of ASD in one person through an oligogenic interaction (Fig. 3b). The underlying genetic interaction may be either additive or by epistasis, although how each CNV influences one's behavior is still not known. Behavioral patterns that determine clinical diagnoses are usually age-dependent. Non-genetic factors may account for the development of symptoms in an age-dependent manner, which may lead to different diagnoses. Taken together, additional conditions, such as interactive secondary genetic loci or any environmental contribution, may influence a person's behavior. Proper medication and early intervention or individualized education[62] may compensate for an individual's particular characteristics derived from the chromosomal variation. A strategic approach combining genetics with individualized treatment may be how ASD is viewed and solved in the future.

ACKNOWLEDGMENTS

Funding for this study was provided by research grants from 'Integrated Research on Neuropsychiatric Disorders' carried out under the Strategic Research Program for Brain Sciences by the Ministry of Education, Culture, Sports, Science and Technology of Japan and The Specific Research Fund 2012 for East Japan Great Earthquake Revival by The New Technology Development Foundation. Grant-in-Aid for Scientific Research on Innovative Areas, 'Glial assembly: A new regulatory machinery of brain function and disorders'.

REFERENCES

1. *Diagnostic and Statistical Manual of Mental Disorders*, 4th edn, text rev. American Psychiatric Association, Washington, 2000.
2. *Diagnostic and Statistical Manual of Mental Disorders*, 5th edn. American Psychiatric Association, Washington, 2013.
3. *ICD-10: International Statistical Classification of Diseases and Related Health Problems*, 10th Rev. edn. World Health Organization, New York, 2008.
4. Freitag CM. The genetics of autistic disorders and its clinical relevance: A review of the literature. *Mol. Psychiatry* 2007; **12**: 2–22.
5. Miyahara M. Meta review of systematic and meta analytic reviews on movement differences, effect of movement based interventions, and the underlying neural mechanisms in autism spectrum disorder. *Front. Integr. Neurosci.* 2013; **7**:16. doi: 10.3389/fnint.2013.00016
6. Wing L. *The Autistic Spectrum: A Guide for Parents and Professionals*. Robinson Publishing, London, 1996.
7. Crespi B, Stead P, Elliot M. Evolution in health and medicine Sackler colloquium: Comparative genomics of autism and schizophrenia. *Proc. Natl Acad. Sci. U.S.A.* 2010; **107** (Suppl. 1): 1736–1741.
8. Kim YS, Leventhal BL, Koh YJ *et al*. Prevalence of autism spectrum disorders in a total population sample. *Am. J. Psychiatry* 2011; **168**: 904–912.
9. Hallmayer J, Cleveland S, Torres A *et al*. Genetic heritability and shared environmental factors among twin pairs with autism. *Arch. Gen. Psychiatry* 2011; **68**: 1095–1102.
10. International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. *Nature* 2004; **431**: 931–945.
11. Sebat J, Lakshmi B, Malhotra D *et al*. Strong association of de novo copy number mutations with autism. *Science* 2007; **316**: 445–449.
12. Kidd JM, Cooper GM, Donahue WF *et al*. Mapping and sequencing of structural variation from eight human genomes. *Nature* 2008; **453**: 56–64.
13. Anney R, Klei L, Pinto D *et al*. A genome-wide scan for common alleles affecting risk for autism. *Hum Mol Genet.* 2010; **19**: 4072–4082.
14. Stein JL, Parikshak NN, Geschwind DH. Rare inherited variation in autism: Beginning to see the forest and a few trees. *Neuron* 2013; **77**: 209–211.

15. Szatmari P, Paterson AD, Zwaigenbaum L *et al*. Mapping autism risk loci using genetic linkage and chromosomal rearrangements. *Nat. Genet.* 2007; **39**: 319–328.
16. Rosenberg LE, Rosenberg DD. *Human Genes and Genomes, Science, Health, Society*. Elsevier Inc., London, 2012.
17. Sebat J, Lakshmi B, Troge J *et al*. Large-scale copy number polymorphism in the human genome. *Science* 2004; **305**: 525–528.
18. Freeman JL, Perry GH, Feuk L *et al*. Copy number variation: New insights in genome diversity. *Genome Res.* 2006; **16**: 949–961.
19. Malhotra D, Sebat J. CNV: Harbingers of a rare variant revolution in psychiatric genetics. *Cell* 2012; **148**: 1223–1241.
20. Ebert DH, Greenberg ME. Activity-dependent neuronal signalling and autism spectrum disorder. *Nature* 2013; **493**: 327–337.
21. Sanders SJ, Ercan-Sencicek AG, Hus V *et al*. Multiple recurrent de novo CNV, including duplications of the 7q11.23 Williams syndrome region, are strongly associated with autism. *Neuron* 2011; **70**: 863–885.
22. Pinto D, Pagnamenta AT, Klei L *et al*. Functional impact of global rare copy number variation in autism spectrum disorders. *Nature* 2010; **466**: 368–372.
23. Devlin B, Scherer SW. Genetic architecture in autism spectrum disorder. *Curr. Opin. Genet. Dev.* 2012; **22**: 229–237.
24. Cooper GM, Coe BP, Girirajan S *et al*. A copy number variation morbidity map of developmental delay. *Nat. Genet.* 2011; **43**: 838–846.
25. Sebat J. Major changes in our DNA lead to major changes in our thinking. *Nat. Genet.* 2007; **39**: S3–S5.
26. Database of Chromosomal Imbalance andPhenotype in Humans Using Ensembl Resources (DECIPHER). Syndromes. 2009. [Cited 5 November 2013.] Available from http://decipher.sanger.ac.uk/syndromes.
27. Horev G, Ellegood J, Lerch JP *et al*. Dosage-dependent phenotypes in models of 16p11.2 lesions found in autism. *Proc. Natl Acad. Sci. U.S.A.* 2011; **108**: 17076–17081.
28. Nakatani J, Tamada K, Hatanaka F *et al*. Abnormal behavior in a chromosome-engineered mouse model for human 15q11-13 duplication seen in autism. *Cell* 2009; **137**: 1235–1246.
29. Liu P, Carvalho CM, Hastings PJ, Lupski JR. Mechanisms for recurrent and complex human genomic rearrangements. *Curr. Opin. Genet. Dev.* 2012; **22**: 211–220.
30. Lupski JR. Genomic rearrangements and sporadic disease. *Nat. Genet.* 2007; **39**: S43–S47.
31. Zhang F, Gu W, Hurles ME, Lupski JR. Copy number variation in human health, disease, and evolution. *Annu. Rev. Genomics Hum. Genet.* 2009; **10**: 451–481.
32. Itsara A, Cooper GM, Baker C *et al*. Population analysis of large copy number variants and hotspots of human genetic disease. *Am. J. Hum. Genet.* 2009; **84**: 148–161.
33. Yoon S, Xuan Z, Makarov V, Ye K, Sebat J. Sensitive and accurate detection of copy number variants using read depth of coverage. *Genome Res.* 2009; **19**: 1586–1592.
34. Schuster SC. Next-generation sequencing transforms today's biology. *Nat. Methods* 2008; **5**: 16–18.
35. Alkan C, Sajjadian S, Eichler EE. Limitations of next-generation genome sequence assembly. *Nat. Methods* 2011; **8**: 61–65.
36. Visscher PM, Brown MA, McCarthy MI, Yang J. Five years of GWAS discovery. *Am. J. Hum. Genet.* 2012; **90**: 7–24.
37. SFARIgene: An evolving database for the autism research community. 2010. [Cited 7 November 2013.] Available from https://gene.sfari.org/autdb/CNVHome.do.
38. Eichler EE, Nickerson DA, Altshuler D *et al*. Completing the map of human genetic variation. *Nature* 2007; **447**: 161–165.
39. Cross S, Kim SJ, Weiss LA *et al*. Molecular genetics of the platelet serotonin system in first-degree relatives of patients with autism. *Neuropsychopharmacology* 2008; **33**: 353–360.
40. Bi W, Park SS, Shaw CJ, Withers MA, Patel PI, Lupski JR. Reciprocal crossovers and a positional preference for strand exchange in recombination events resulting in deletion or duplication of chromosome 17p11.2. *Am. J. Hum. Genet.* 2003; **73**: 1302–1315.
41. Lupski JR. Retrotransposition and structural variation in the human genome. *Cell* 2010; **141**: 1110–1112.
42. Weiss LA. Autism genetics: Emerging data from genome-wide copy-number and single nucleotide polymorphism scans. *Expert. Rev. Mol. Diagn.* 2009; **9**: 795–803.
43. Sutcliffe JS. Genetics: Insights into the pathogenesis of autism. *Science* 2008; **321**: 208–209.
44. Glessner JT, Wang K, Cai G *et al*. Autism genome-wide copy number variation reveals ubiquitin and neuronal genes. *Nature* 2009; **459**: 569–573.
45. Tabuchi K, Blundell J, Etherton MR *et al*. A neuroligin-3 mutation implicated in autism increases inhibitory synaptic transmission in mice. *Science* 2007; **318**: 71–76.
46. Yashiro K, Riday TT, Condon KH *et al*. Ube3a is required for experience-dependent maturation of the neocortex. *Nat. Neurosci.* 2009; **12**: 777–783.
47. Wallace ML, Burette AC, Weinberg RJ, Philpot BD. Maternal loss of Ube3a produces an excitatory/inhibitory imbalance through neuron type-specific synaptic defects. *Neuron* 2012; **74**: 793–800.
48. Girirajan S, Johnson RL, Tassone F *et al*. Global increases in both common and rare copy number load associated with autism. *Hum. Mol. Genet.* 2013; **22**: 2870–2880.
49. Gilman SR, Chang J, Xu B *et al*. Diverse types of genetic variation converge on functional gene networks involved in schizophrenia. *Nat. Neurosci.* 2012; **15**: 1723–1728.
50. Simons Variation in Individuals Project (Simons VIP): A genetics-first approach to studying autism spectrum and related neurodevelopmental disorders. *Neuron* 2012; **73**: 1063–1067.

51. Zufferey F, Sherr EH, Beckmann ND *et al*. A 600 kb deletion syndrome at 16p11.2 leads to energy imbalance and neuropsychiatric disorders. *J. Med. Genet.* 2012; **49**: 660–668.
52. Michaelovsky E, Frisch A, Carmel M *et al*. Genotype-phenotype correlation in 22q11.2 deletion syndrome. *BMC Med. Genet.* 2012; **13**: 122.
53. Bassett AS, Scherer SW, Brzustowicz LM. Copy number variations in schizophrenia: Critical review and new perspectives on concepts of genetics and disease. *Am. J. Psychiatry* 2010; **167**: 899–914.
54. Vorstman JA, Morcus ME, Duijff SN *et al*. The 22q11.2 deletion in children: High rate of autistic disorders and early onset of psychotic symptoms. *J. Am. Acad. Child Adolesc. Psychiatry* 2006; **45**: 1104–1113.
55. Niklasson L, Rasmussen P, Oskarsdottir S, Gillberg C. Neuropsychiatric disorders in the 22q11 deletion syndrome. *Genet. Med.* 2001; **3**: 79–84.
56. Fine SE, Weissman A, Gerdes M *et al*. Autism spectrum disorders and symptoms in children with molecularly confirmed 22q11.2 deletion syndrome. *J. Autism Dev. Disord.* 2005; **35**: 461–470.
57. Girirajan S, Rosenfeld JA, Cooper GM *et al*. A recurrent 16p12.1 microdeletion supports a two-hit model for severe developmental delay. *Nat. Genet.* 2010; **42**: 203–209.
58. Guruharsha KG, Kankel MW, Artavanis-Tsakonas S. The Notch signalling system: Recent insights into the complexity of a conserved pathway. *Nat. Rev. Genet.* 2012; **13**: 654–666.
59. Louvi A, Artavanis-Tsakonas S. Notch signalling in vertebrate neural development. *Nat. Rev. Neurosci.* 2006; **7**: 93–102.
60. Nagaraj R, Banerjee U. The little R cell that could. *Int. J. Dev. Biol.* 2004; **48**: 755–760.
61. Simon MA, Bowtell DD, Dodson GS, Laverty TR, Rubin GM. Ras1 and a putative guanine nucleotide exchange factor perform crucial steps in signaling by the sevenless protein tyrosine kinase. *Cell* 1991; **67**: 701–716.
62. Warren Z, McPheeters ML, Sathe N, Foss-Feig JH, Glasser A, Veenstra-Vanderweele J. A systematic review of early intensive intervention for autism spectrum disorders. *Pediatrics* 2011; **127**: e1303–e1311.