We are IntechOpen,
the world's leading publisher of
Open Access books
Built by scientists, for scientists

5,300
Open access books available

129,000
International authors and editors

155M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS

BOOK CITATION INDEX

INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

Genetic Evaluation of Individuals with Autism Spectrum Disorders

Eric C. Larsen, Catherine Croft Swanwick and
Sharmila Banerjee-Basu

Additional information is available at the end of the chapter

1. Introduction

While the genetic component of Autism Spectrum Disorders (ASD) has been clearly established from various lines of study, the multitude of genes and chromosomal loci associated with ASD has made identification of the underlying molecular mechanisms of pathogenesis difficult to resolve. A range of diverse methodologies and study types have identified both rare and common genetic variants in ASD candidate genes and chromosomal loci. Moreover, the recent development of high-throughput next generation sequencing (NGS) technologies and the increasing usage of chromosomal microarray analysis (CMA) has led to a significant expansion in the number of single nucleotide variants (SNVs) and copy number variants (CNVs) potentially affecting one or more genes that have been identified in ASD individuals. This, in turn, has given critical insight into the molecular and cellular processes that may be preferentially targeted for disruption by genetic lesions in ASD patients.

However, it is important to note that there is no genetic test available for the diagnosis of ASD. Rather, genetic testing is primarily aimed at identifying genetic variants potentially responsible for disease pathogenesis in a given individual diagnosed with ASD. Furthermore, the utility of NGS and CMA in genetic evaluation of ASD individuals is dependent on proper interpretation and reporting of test results. In this chapter we will discuss 1) genetic testing technologies currently available for the identification of genetic variation in ASD cases, 2) the genes and genomic loci targeted by single nucleotide and copy number variants that have been linked to ASD susceptibility, 3) the bioinformatics tools that enable researchers to process the enormous amount of genetic data associated with ASD, and 4) challenges that exist in the interpretation and reporting of genetic evaluation results in ASD cases.

2. Genetic screening technologies for the evaluation of ASD cases

Autism spectrum disorders (ASD) are among the most highly heritable neurodevelopmental disorders, and extensive research has been focused on identifying the underlying genetic basis of these disorders. It has become apparent that ASD is a genetically heterogeneous disorder, with hundreds of genes and chromosomal rearrangements identified that confer varying degrees of risk for disease. Initially, susceptibility genes and genomic loci were identified by costly, low-throughput techniques, such as automated Sanger sequencing and conventional cytogenetic techniques. The need for lower-cost, higher-throughput genetic screening technologies capable of identifying genome-wide variation in individuals with genetically complex diseases, such as ASD, has driven improvements in pre-existing techniques and the development of new technologies. The genetic screening technologies presently available to clinical geneticists and researchers are capable of providing lower-cost, high-throughput genetic data that have significantly expanded our knowledge of genetic variation, both in the general population and in ASD individuals in particular. The first major high-throughput studies aimed at identifying CNVs and SNVs in ASD cohorts were published in 2010 and 2011, respectively [1, 2]. Here we describe in greater detail two of these genetic screening technologies that have become widely used in the genetic evaluation of ASD cases: next generation sequencing (NGS) and chromosomal microarray (CMA).

2.1. Next generation sequencing

Next-generation sequencing (NGS) is a term used to describe a collection of high-throughput sequencing technologies that have enabled clinicans to screen larger amounts of genetic material at lower cost than traditional sequencing technologies, such as automated Sanger sequencing [3]. NGS is typically used to identify single nucleotide variants (SNVs), as well as small insertions or deletions in candidate genes. However, NGS can also be used to identify copy number variants (CNVs), as was recently demonstrated in a report detailing whole exome sequencing in a cohort of ASD cases [4], as well as balanced chromosomal rearrangments, which are typically not detected by genome-wide microarrays [5, 6]. Since 2011, six research articles have been published that have identified rare variants in both existing and novel ASD susceptibility genes using NGS techniques [2, 4, 7-10], a fact that illustrates how extensively these techniques have been adopted by the ASD research community. As a result of these studies, the potential number of potential ASD-linked genes have increased dramatically (Table 1). Furthermore, as demonstrated by the minimal overlap of candidate genes across these studies, the results of these studies further illustrate the genetic heterogeneity of ASD.

NGS techniques are typically divided into three categories, each with its own advantages and disadvantages (summarized in Table 2). These techniques vary in terms of genetic coverage (the size of the sequenced target, which can range in size from one or a few genes to the entire genome) and genetic resolution (sensitivity in detection of variants per sequencing target). In general, the smaller the genetic coverage, the higher the genetic resolution. It should be noted that, as the size of the target for sequencing increases, so does the number

of both false-positive and false-negative variants [11]; this is one of the many considerations that must be taken into account in deciding which NGS technique to utilize.

Exome Report	Total # of genes	# of unique genes	# of overlapping genes
O'Roak 2011	21	21	0
Sanders 2012	170	166	4
O'Roak 2012	240	227	13
Neale 2012	173	168	5
Chahrour 2012	53	53	0
Iossifov 2012	363	338	25
Total	1020	973	47

Table 1. Six recent scientific articles describing rare genetic variants identified in ASD cases by next generation sequencing (NGS) has led to a dramatic increase in the number of potential ASD candidate genes and further illustrates the genetic heterogeneity of ASD. Overlapping genes are genes in which a rare variant was identified in more than one exome report and are used as a measure of genetic heterogeneity.

	Availability	Cost	Advantages	Disadvantages
Targeted gene panels	Commercially available	~$5000	-Highest resolution of NGS approaches -Contains a number of well-characterized ASD susceptibility genes (syndromic and non-syndromic)	-Unable to detect variants in genes not included in the panel
Whole exome sequencing	Available only in research settings	~$1000	-Estimated to detect the majority of disease-covering variants	-Unable to detect variants in non-coding regions
Whole genome sequencing	Available only in research settings	~$4000-$5000	-Greatest coverage of the genome (both coding and non-coding regions)	-Lowest resolution of NGS techniques

	Availability	Cost	Advantages	Disadvantages
				-Higher cost than whole exome sequencing

Table 2. A summary of the benefits and drawbacks of the three types of next generation sequencing (NGS) techniques in the genetic evaluation of ASD cases. Cost estimates of whole-exome and whole-genome sequencing in [14].

2.1.1. Targeted gene panels

Targeted gene panels generally test for 50-100 genes that have been demonstrated to be strongly associated with a particular disease. Such gene panels are already extensively used to screen individuals for a wide range of cancers and inherited diseases for which causative genes have been identified. A number of commercially-available ASD gene panels have recently been designed to target both genes strongly associated with non-syndromic ASD as well as syndromic genes (genes that cause syndromes in which a subset of affected individuals also develop ASD, such as *FMR1*, *MECP2*, and *CACNA1C*, which cause Fragile X, Rett, and Timothy syndromes, respectively). For example, the Greenwood Genetic Center offers a 62-gene syndromic gene panel that covers the coding region and flanking intronic boundaries of ASD-linked 62 genes for $5500 (http://www.ggc.org/images/pdfs/syndromicautism62-genengspanel.pdf). While targeted gene panels offer the smallest coverage of the human genome of the three NGS approaches, they offer the highest resolution. One of the major drawbacks to the use of targeted gene panels for a genetically heterogeneous disorder such as ASD is the inability to detect mutations in genes outside of those included in the gene panel.

2.1.2. Whole exome sequencing

Whole exome sequencing, which is also known as targeted exome capture, is designed to specifically identify variants in protein-coding regions of the human genome. Although these protein-coding regions, called exons, constitute a very small percentage of the human genome, it is estimated that they contain up to 85% of disease-causing mutations [12, 13]. However, whole exome sequencing will fail to detect any potentially pathogenic variants in non-coding regions of the human genome. This NGS method also provides lower resolution than targeted gene panels. Nonetheless, whole exome sequencing is increasingly being used to identify potentially pathogenic rare single gene variants in individuals with ASD [2, 4, 7-10].

2.1.3. Whole genome sequencing

In contrast to whole exome sequencing, which only covers protein-coding regions of the human genome, whole genome sequencing provide coverage of the entire genome, allowing for the sequencing of both coding and non-coding genomic regions. As such, single nucleotide changes and small insertions/deletions within non-coding regions of the genome can be detected by this method. While whole genome sequencing covers the largest amount of the human genome of all NGS techniques, it offers the lowest resolution of the three NGS technologies. Whole genome sequencing is also more costly than whole exome sequencing,

although the differences in cost betwen these two techniques have fallen from 10- to 20-fold [13] to 4- to 5-fold [14].
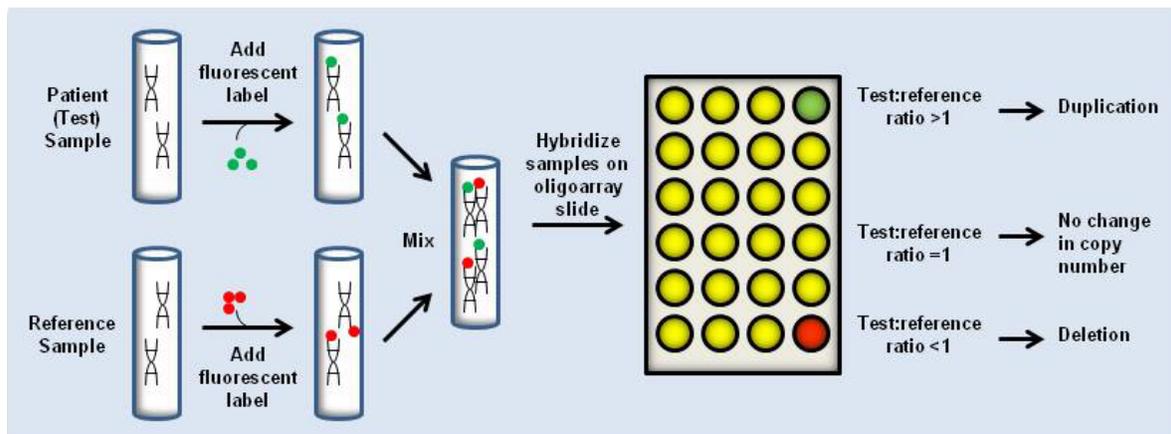
2.2. Chromosomal microarray

Microscopically-visible chromosomal rearrangments have long been implicated in the onset and pathogenesis of neurodevelopmental disorders, includng ASD. Indeed, many of the most strongly ASD-linked chromosomal deletions and duplications, collectively referred to as copy number variants (CNVs), were discovered through the use of conventional cytogenetic techniques such as G-banded karyotyping, fluorescent in situ hybridization (FISH), and microsatellite analysis. For example, duplications of chromosome 15q11-q13 were first implicated in ASD in the mid-1990s by these methods [15-17]. Likewise, these methods identified chromosomal rearrangments on the long arm of chromosome 22 in ASD cases [18, 19]. However, conventional cytogenetic techniques are impractical in the identification of copy number variation throughout the human genome in large case cohorts. While G-banded karyotyping is capable of detecting large chromosomal deletions and duplications (~1 Mb and larger), it lacks the sensitivity to detect smaller CNVs. Alternatively, the use of techniques such as FISH is generally limited to screen a particular chromosomal region, so while they are useful for examining copy number variation in a genomic loci of interest in larger case populations, they are impractical for the purposes of identifying deletions and duplications throughout the genome.

In the last decade, technological and computational advances have allowed clinical geneticists and researchers to detect submicroscopic chromosomal deletions and duplications throughout the human genome in large case cohorts that would not be detected by traditional cytogenetic techniques. Chromosomal microarray (CMA) is a term frequently used to include all types of array-based whole genome copy number analyses, with the two most widely used being array-comparative genomic hybridization (aCGH) and single nucleotide polymorphism (SNP) arrays. CMA has been demonstrated to provide a higher diagnostic yield than G-banded karyotyping (15-20% compared to ~3%) due to its ability to detect submicroscopic deletions and duplications, and it has been proposed that CMA should replace conventional cytogenetic techniques as a first-tier diagnostic tool for individuals with congential abnormalities and developmental disorders, including ASD [20]. High-throughput genome-wide aCGH and SNP arrays are now regularly used in the detection of CNVs in large ASD cohorts [1, 21-24].

aCGH and SNP arrays employ similar methodologies in the detection of CNVs (Figure 1). The first step involves labeling the DNA of the ASD patient with a fluorophore, thereby creating a test sample. The test sample is then mixed with an equal amount of DNA from a normal reference sample that has been labeled with a different fluorophore. This mixed DNA sample is added to a glass slide containing thousands of oligonucleotide probes corresponding to different chromosomal regions that cover the human genome; in the case of SNP arrays, the oligonucleotide probes are specific for common polymorphisms found in the general population. The sensitivity of CMA has been greatly increased in recent years by the development of arrays employing a larger number of smaller oligonulceotide probes; in doing so, clinical geneticists and researchers are able to detect even smaller copy number changes than before

without compromising genomic coverage. The test and reference DNA samples hybridize with the probes on the slide, and the fluorescence intensities of the test and reference DNA can then be measured. Following analysis with software that is typically specific for the platform being used, one or more algorithms are used to call the CNV. The ratio between the two fluorescence intensities is used to identify copy number changes. For example, if the test-to-reference ratio is 1 (yellow in the example below), then there is no change in copy number at the chromosomal region corresponding to a given probe, If the test-to-reference fluorescence ratio is > 1 for a particular probe (green in the example below), then the ASD patient carries a duplication in the chromosomal region corresponding to that probe. If the test-to-reference ratio is < 1 (red in the example below), then the patient carries a deletion at that site of the genome.


Figure 1. Chromosomal microarray (CMA) analysis involves hybridization of differently labeled test and reference DNA samples with oligonucleotide probes, followed by computerized analysis and identification of copy number variants (CNVs) based on changes in fluorescence intensity ratio.

Despite the recommended use of CMA as a first-tier genetic evaluation tool in place of conventional cytogenetic techniques, it should be noted that aCGH is unable to detect balanced chromosomal rearrangments and other chromosomal abnormalities that have traditionally been detected by karyotype analysis [25]. In addition to their traditional utilization in the detection of risk-conferring common polymorphisms, SNP arrays have the added advantage of being able to detect copy number neutral genetic variation such as uniparental disomy and long contiguous streteches of homozygosity (LCSH) that cannot be detected by aCGH [25, 26].

3. Genetic variation in ASD

With the advent of NGS techniques and increasing usage of CMA screening, the number of SNVs and CNVs that have been identified in ASD individuals has grown significantly. Based on a survey of recently published exome sequencing studies of ASD cohorts, it was estimated that the number of dosage-sensitive ASD susceptibility genes is approximately 370, with roughly a third of these genes having been identified [10]. However, even this number might be a conservative projection. As shown in Figure 2, the number of ASD susceptibility genes in

the Human Gene Module of the autism genetic database AutDB [27] has increased from 284 genes in September 2011 to 369 genes in June 2012. A large number of newer susceptibility genes have been annotated from reports employing whole exome sequencing of ASD cases [2, 4, 7-10], illustrating the increasing usage of NGS techniques in the study of genetic variation in ASD. In addition to the identification of novel ASD susceptibility genes, NGS techniques have identified novel rare variants in previously identified ASD susceptibility genes. The number of ASD-associated CNV loci has also increased significantly, with the CNV module of AutDB expanding from 1034 CNV loci in September 2011 to 1173 loci in June 2012 (Figure 2). In this section we describe the genetic categories into which ASD susceptibility genes have been classified, as well as describe recent studies that have yielded invaluable insight on the functional profiles of ASD-associated genes and CNV loci.
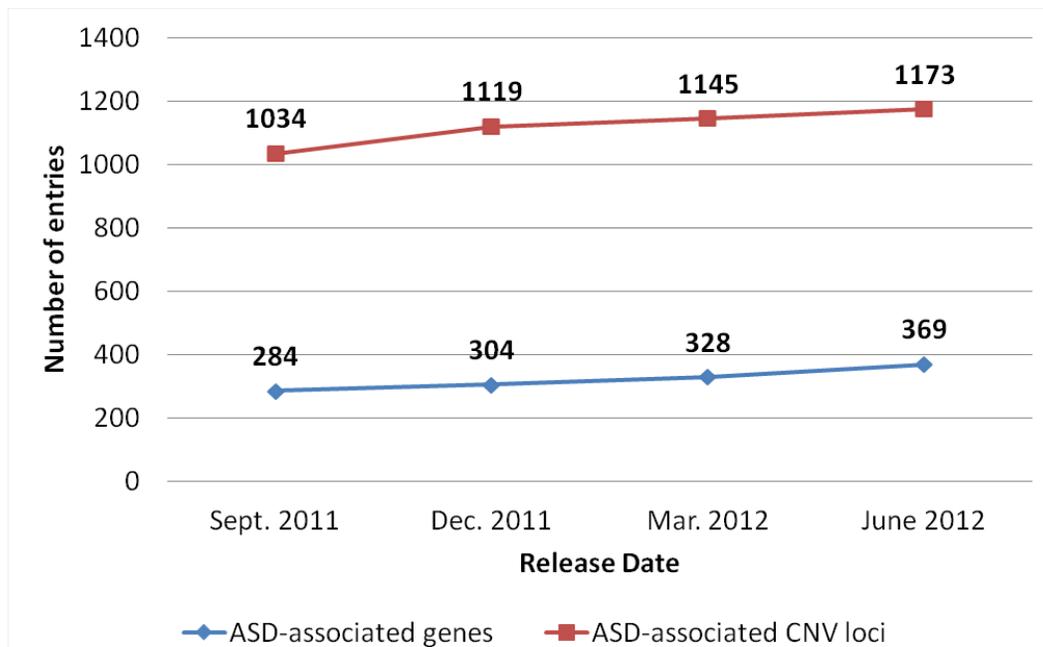
Figure 2. The number of genes and CNV loci associated with ASD in the genetic database AutDB has increased over the last four quarterly release dates.

3.1. Genetic categories of ASD susceptibility genes

The earliest ASD susceptibility genes were rare single gene variants in genes associated with syndromes such as Fragile X syndrome and Rett syndrome. The discovery of single gene mutations/disruptions in two neuroligin genes, NLGN3 and NLGN4, in ASD siblings [28] initiated the search for additional ASD susceptibility genes in non-syndromic ASD cases. The continued identification of rare genetic variants associated with both syndromic and non-syndromic ASD, as well as of risk-conferring polymorphisms enriched in ASD populations

compared to unaffected controls in genetic association studies, has led to significant increases in the number of ASD-linked genes. While the majority of ASD-associated genes have been linked to disease on the basis of genetic studies in human populations, a number of additional ASD-linked genes have been identified by alternate methodologies, such as gene expression studies in post-mortem brain tissue of ASD individuals.

ASD susceptibilty genes in the Human Gene Module of AutDB are defined into four distinct categories:

1. Rare. This category features genes implicated in rare monogenic forms of non-syndromic ASD. Rare allelic variants within this category include single nucleotide variants, small insertions and deletions, chromosomal rearrangements such as translocations and inversions, and monogenic submicroscopic deletions and duplications. Among the genes within this category are *CACNA1H* and *SHANK1.*

2. Syndromic. Syndromic genes were among the first genes for which rare genetic variants linked to autism were identifed. In addition to well-characterized syndromic genes such as *FMR1* (Fragile X syndrome), *MECP2* (Rett syndrome), and *CACNA1C* (Timothy syndrome), genes such as *CHD7* and *SLC9A6* fall into the syndromic category.

3. Association. This category includes genes in which small risk-conferring common poly-morphisms have been identified from genetic association studies in idiopathic ASD populations. Among the genes within this category are *MET* and *MTHFR.*

4. Functional. This category includes functional candidate genes that have not yet been experimentally linked to ASD by genetic studies. Among the genes in this category are *BCL2* and *PDE4B*, whose inclusion is based on changes in gene expression in post-mortem brain tissue of ASD subjects.

As shown in Figure 3, while the number of both rare and common ASD-associated variants in the Human Gene module of AutDB has increased over the last four quarterly release dates, the number of rare variants has increased at a much greater rate than the number of common variants. The number of rare variants increased from 1141 in September 2011 to 1675 in June 2012, an increase of ~146%. In contrast, the number of common variants rose form 508 to 575 over the same span of time, an increase of only ~113%. This disparity between the addition of rare and common variants to AutDB is in part due to the increased usage of NGS and CMA and subsequent identification of rare ASD-associated variants in large ASD cohorts.

It should be noted that a given gene can fall under multiple genetic categories, depending on the affected population under investigation and the type of study. For example, both rare variants and risk-conferring common polymorphisms have been identified in the *CNTNAP2* gene in ASD individuals across multiple studies [2, 29-31] However, in addition to its role as an ASD susceptibility factor, recent studies suggest that rare variants in *CNTNAP2* are responsible for two additional syndromes: cortical dysplasia-focal epilepsy syndrome [32] and Pitt-Hopkins-like syndrome 1 [33]. Therefore, based on the combined evidence from all of these aforementioned studies, *CNTNAP2* is classified in AutDB as a syndromic gene, a rare gene, and an association gene.

The classification of ASD-linked genes into genetic categories is a useful tool in assessing the strength of the evidence for the connection of a given gene with ASD. Genes within the rare and syndromic categories are generally considered to have the strongest link to ASD [34]. Due to the frequent lack of replication in their association with ASD from one study to the next, genes within the association category are considered to have a weaker link to ASD than genes within the rare and syndromic categories. Genes within the functional category have no direct documented connection to ASD and are therefore considered to be among the weakest ASD candidate genes.
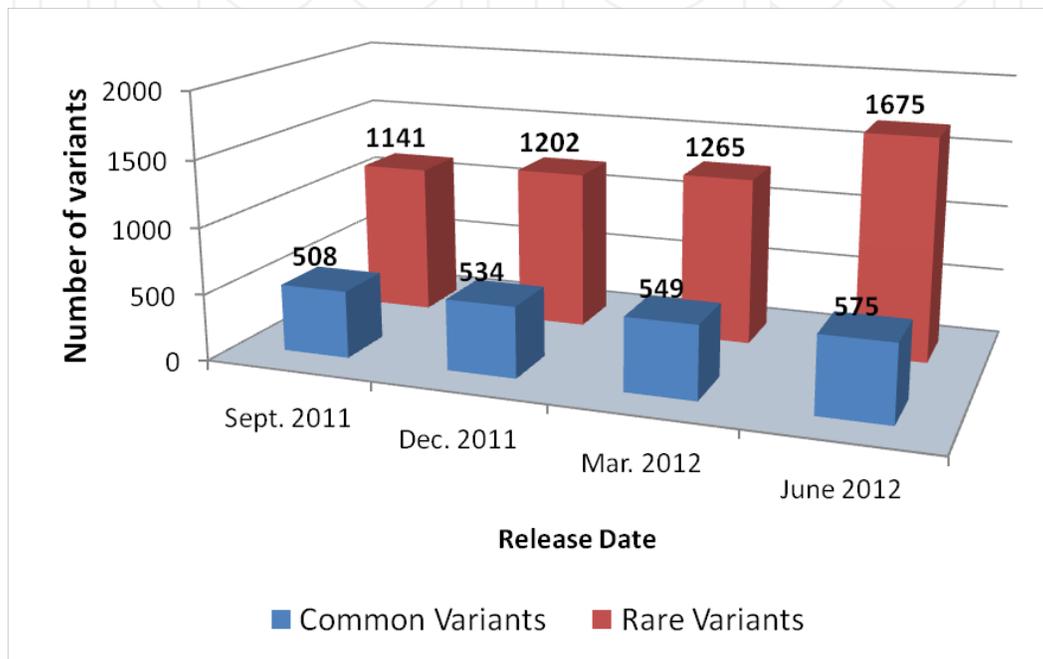
Figure 3. The number of rare and common ASD-associated variants in the Human Gene module of AutDB has increased over the last four quarterly release dates.

3.2. Functional profiles of ASD-associated genes and CNV loci

The increasing number of ASD-associated genetic factors, as shown in Figure 2, has only added to the well-established genetic heterogeneity of ASD. In spite of the complexity caused by this genetic heterogeneity, bioinformatic analysis of ASD-linked genes and CNV loci has yielded valuable insight into the molecular interactions and cellular pathways preferentially targeted by genetic lesions in individuals with ASD. Not only can this information be potentially used to design therapeutic approaches targeting disrupted pathways, but it can also aid in assessing the clinical importance of newly-discovered ASD candidate genes and CNV loci in which pathogenic variants are identified by NGS and CMA, respectively. For example, a gene whose encoded gene product resides within a known ASD-associated cellular process or interacts with a known ASD-associated gene is a stronger candidate than a gene that fails to reside within known ASD-associated cellular processes or interact with known ASD-associated genes.

Recent large-scale ASD genetic studies have used a systems biology approach to translate genetic information into functional profiles that shed light on how genetic variation in ASD may lead to disease onset and pathogenesis. Rare CNVs identified in large ASD cohorts have been shown to be enriched for genes involved in cellular processes of relevance for ASD, including cellular proliferation, projection, and motility, and GTPase/Ras signaling [1], neuronal cell adhesion and ubiquitin-mediated degradation [22], glycobiology [35], axon growth and pathfinding [36], and synapse development, axon targeting, and neuron motility [37]. Gene datasets from genome-wide association studies in ASD populations were demonstrated to be enriched for Gene Ontology (GO) classifications for cellular processes including pyruvate metabolism, transcription factor activation, cell signaling and cell-cycle regulation [38]. A recent report describing gene pathway analysis using single nucleotide polymorphism (SNP) data from the Autism Genetics Research Exchange (AGRE) identified cellular pathways such as calcium signaling, long-term depression and potentiation, and phosphotidylinositol signaling that reached statistical significance in both Central European and Han Chinese populations [39]. More recently, whole exome sequencing studies in large ASD cohorts have demonstrated that proteins encoded by genes in which potentially disruptive *de novo* mutations were identified showed a higher degree of connectivity among themselves and to previously identified ASD genes based on protein-protein interaction network analysis [4, 8]. Another exome sequencing study in ASD individuals found that many of the genes in which potentially disruptive variants were identified associated with the Fragile X Mental Retardation Protein (FMRP), the encoded product of the syndromic ASD gene *FMR1* [10]. Taken together, these functional maps suggest that specific cellular pathways and processes are preferentially targeted by genetic variation in ASD cases, and that association with the encoded products of well-characterized ASD-linked genes offers evidence for pathogenic relevance.

Knowledge of ASD-associated genes can also be used to identify novel ASD candidate genes. Following the construction of functional and expression profiles from a reference set of 84 rare and syndromic ASD-linked genes, we generated a predictive map of novel ASD candidate genes [40]. In total, 460 potential candidate genes were identified that overlapped both the functional profile and the brain expression profile of the initial reference set. The power of this predictive gene map was demonstrated by the capture of 18 pre-existing ASD-associated genes that were not included in the reference gene dataset, with the remaining 442 genes serving as novel ASD candidate genes. Since the publication of our predictive gene map, 12 of the novel ASD candidate genes identified in [40] have been added to AutDB, demonstrating the continued power of this analysis (manuscript in preparation).

4. Bioinformatics of ASD

With the rapid growth of genetic data obtained from ASD individuals, there has become a critical need for databases specializing in the storage and assessment of this data. Here we highlight several of the ASD-related genetics databases that are available to researchers.

4.1. AutDB

Our autism database AutDB (http://autism.mindspec.org/autdb/Welcome.do) is a web-based, searchable database of ASD candidate genes identified in genetic association studies, genes linked to syndromic autism, and rare single gene mutations [27]. Evidence regarding ASD candidate genes is systematically extracted from peer-reviewed, primary scientific literature and manually curated by our researchers for inclusion in AutDB. To provide high-resolution view of various components linked to ASD, we developed detailed annotation rules based on the biology of each data type and generated controlled vocabulary for data representation. AutDB is widely used by individual laboratories in the ASD research community, as well as by consortiums such as the Simons Foundation, which licenses it as SFARI Gene.

AutDB is designed with a systems biology approach, integrating genetic information within the original Human Gene module to corresponding data in subsequent Animal Model, Protein Interaction (PIN) and Copy Number Variant (CNV) modules. The Animal Model module contains a comprehensive collection of mouse models linked to ASD [41]. While the Animal Model module initially contained only genetic mouse models of ASD, it has since been expanded to include induced mouse models of ASD in which a chemical or biological agent linked to ASD has been administered. As core behavioral features of ASD such as social interactions and communications can only be approximated in animal models, the annotation strategy for this module includes four broad areas: 1) core behavioral features of ASD, 2) ASD-related traits such as seizures and circadian rhythms that are heritable and more easily quantified in animal models; 3) neuroanatomical features, and 4) molecular profiles. To this end, we developed PhenoBase, a classification table for systematically annotating models with controlled vocabulary containing 16 major categories and >100 standardized phenotype terms. The PIN module of AutDB serves as a repository for all known protein interactions of ASD candidate genes, documenting six major types of direct interactions: 1) protein binding, 2) promoter binding, 3), RNA binding, 4) protein modification, 5) direct regulation, and 6) autoregulation. Its content is envisioned to have immediate application for network biology analysis of molecular pathways involved in ASD pathogenesis. For the purposes of genetic evaluation of individuals with ASD, knowledge of the protein interactions of ASD-associated genes can potentially aid in the clinical assessment of novel ASD candidate genes based on their interactions, or lack thereof, with known ASD-linked genes.

4.2. Gene scoring module of SFARI gene

As previously mentioned, AutDB is licensed to the Simons Foundation as SFARI Gene. However, unlike AutDB, SFARI Gene includes a unique feature initiated by the Simons Foundation called the Gene Scoring module (https://gene.sfari.org/autdb/GS_Home.do). The Gene Scoring module is a web-based platform detailing the rank of ASD-associated genes in the SFARI Gene Human Gene module [42]. With the increase in the number of genes linked to ASD, a Gene Scoring initiative was launched to assess the ASD candidate genes based on a set of standardized annotation rules. Following evaluation by an expert panel of advisors, the

gene assessment results are then integrated in the form of Gene Score Cards to display the scores and the evidence in a graphical user interface for the ASD-linked gene. Recently, a community-wide annotation functionality was incorporated into the Gene Scoring module, allowing users to download the Gene Scoring dataset, score genes of their choice, and submit their scores to SFARI for possible inclusion.

4.3. DECIPHER

DECIPHER (Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources) (http://decipher.sanger.ac.uk/) is an interactive web-based database that incorporates a suite of tools designed to aid in the interpretation of submicroscopic chromosomal deletions and duplications [43]. Genetic and phenotypic information is publically available not only for individuals diagnosed with idiopathic ASD, but also for individuals diagnosed with a recognized microdeletion or microduplication syndrome in which a subset of affected individuals also develop ASD.

4.4. AutismKB

AutismKB (http://autismkb.cbi.pku.edu.cn/) is a web-based, searchable database hosted by the Center for Bioinformatics, Peking University [44]. AutismKB is an evidence-based knowledge resource for ASD genetics containing information on genes, copy number variants, and linkage regions associated with ASD. Analysis of the gene content in AutismKB is available for users in the form of GO term enrichment analysis using the DAVID functional annotation tool and pathway enrichment analysis. Much like the Gene Scoring Module of SFARI Gene (see section 4.2), the genes within AutismKB are scored.

4.5. Autism Chromosome Rearrangement Database

The Autism Chromosome Rearrangement Database (http://projects.tcag.ca/autism/) is a web-based, searchable genetic database of chromosomal structural variation in ASD that is hosted by The Centre for Applied Genomics at the Hospital for Sick Children in Toronto, Canada [21]. The content of this database, which is derived both from published research articles and in-house experimental results, includes cytogenetic and microarray data from individuals with ASD.

4.6. Autism Genetic Database

The Autism Genetic Database (http://wren.bcf.ku.edu/) is a web-based, searchable genetic database developed by researchers at the University of Kansas [45]. In addition to ASD-associated genes and CNVs, this database also includes information on known non-coding RNAs and chemically-induced fragile sites in the human genome.

Recent lines of evidence have placed non-coding RNAs under increased scrutiny with regards to their potential pathogenic role in ASD. A number of small nucleolar RNAs (snoRNAs) reside within the ASD-associated 15q11-q13 region. A mouse model engineered to mimic duplication of the 15q11-q13 region observed in ~1% of ASD cases exhibited overexpression of the snoRNA

MBII52 (the mouse ortholog of the human snoRNA HBII52), which could potentially alter serotonergic signaling and contribute in part to the ASD-associated traits exhibited by these mice [46]. More recently, it was discovered that a non-coding RNA is transcribed from a gene-poor region of chromosome 5p14.1 identified in genome-wide association studies of ASD cohorts [47]. Expression of the non-coding RNA, designated MSNP1AS, was shown to be higher both in individuals carrying the ASD-associated T allele and in post-mortem brain tissue of individuals with ASD.

Spontaneous breakage during DNA replication at rare chromosomal fragile sites may also play a role in the pathogenesis of neuropsychiatric disorders such as ASD. The chromosomal fragile site FRAXA has been implicated in fragile X syndrome, and other fragile sites have been identified that associate with ASD, such as FRA2B, FRA6A, and FRA13A [48].

5. Challenges of genetic evaluation in ASD

NGS and CMA have expanded the ability of clinical geneticists and researchers to identify potential genetic causes of ASD. However, there are many challenges still present in the field of genetic evaluation. A recent report in the *American Journal of Medical Genetics* found that many children with ASD fail to get genetic evaluation, and that parents and medical professionals need to be better educated about the potential benefits of genetic evaluation [49]. Educating parents on genetic evaluation is especially critical in light of a recent survey of nine parents regarding their child's participation in genetic research in ASD [50], in which parents valued having had their child enrolled for a variety of reasons, including the potential use of genetic results in tailoring intervention and in family planning, the establishment of connections with experts in the field of ASD, and networking with other families, among others.

Even with the increased sensitivity of genetic evaluation techniques, an underlying genetic cause of ASD is still only identified in a minority (< 25%) of ASD cases [51]. One of the major challenges in the clinical interpretation of NGS and CMA lies in differentiating between pathogenic and benign genetic variants identified in ASD patients. The pathogenic relevance of the vast majority of ASD-linked genetic variants remains unknown; such variants are frequently classified as variants of unknown significance, or VOUS. While the identification of a genetic lesion in an existing ASD susceptibility gene or CNV locus is suggestive of a possible genetic cause of disease, variants in these genes and CNV loci have also been observed in seemingly unaffected individuals. Furthermore, it is important to note that, while technological advances have expanded the ability of clinical geneticists and researchers to identify these potential genetic causes of ASD, there is no genetic test available for the diagnosis of ASD. A recent report proposed a means of predicting a diagnosis of ASD based on the identification of candidate SNPs [39]. The accuracy of the predictive classifier was found to be 71.7% in individuals of Central European descent from validation datasets. However, the accuracy of the predictive classifier fell when tested in a Han Chinese cohort, a finding that stresses how genetic heterogeneity across populations complicates the use of such an ap-

proach. In addition, the overall accuracy of the predictive classifier is likely too low to serve as an effective diagnostic tool.

A number of guidelines have already been proposed to aid clinicians and clinical geneticists in the interpretation and reporting of CNVs. With the increasing use of high resolution NGS technologies, similar guidelines will likely be proposed for the interpretation and reporting of single nucleotide variants (SNVs). Furthermore, tools and prioritization schema have also been developed to aid clinicians in the interpretation of genetic testing results. Here we discuss in greater detail the challenges in interpreting genetic screening results in ASD cases, the strategies that have been proposed for the interpretation and reporting of screening results, and the resources available to aid in that interpretation.

5.1. Challenges in the interpretation of ASD genetic screening results

5.1.1. Technical limitations of NGS and CMA

As previously mentioned, as the size of the sequenced target increases, so does the potential number of false-positive and false-negative variants identified [11]. Such sequencing artifacts are particularly problematic for the detection of spontaneous, or *de novo* variants, as false-positive variants would appear to be *de novo* in origin when they are observed in an offspring's genome but not in parental genomes. Furthermore, the source of DNA used in sequencing studies can introduce sequencing artifacts. DNA from lymphoblastoid cell lines from individuals to be genetically evaluated is a commonly used template for sequencing; however, the creation and culturing of these cell lines can introduce genetic changes that would appear as *de novo* variants when such cell lines are compared between parents and offspring. In order to remove or reduce the possibility of artifactual results, subsequent variant validation should be performed. In the case of single gene variants identified by NGS, a more targeted sequencing approach limited to the gene or region of interest would confirm the variant previously identified. In the case of CNVs identified by NGS or CMA, a targeted detection method such as quantitative real-time PCR or FISH is frequently used to confirm their discovery.

5.1.2. Genetic heterogeneity

While the genetic basis of many human diseases can be traced back to one or a few genes, the genetic basis of complex neuropsychiatric disorders such as ASD has proven to be far more complicated, with hundreds of genes and genomic loci associated with varying risks of disease. The recent utilization of NGS and CMA approaches in genetic evaluation of ASD cases has led to the detection of genetic variation not only in both existing and novel susceptibility genes and genomic loci. However, the strength of evidence for many of these novel candidate genes or genomic loci is minimal, and some degree of replication in follow-up studies will be required to fully assess the relevance of many of these newly-identified variants.

5.1.3. Incompelete penetrance and variable expressivity

One of the major challenges in identifying potential causative genetic variation in ASD cases lies in the fact that a potentially disruptive variant in a gene or genomic loci may not always associate or segregate with disease. For example, a potentially pathogenic variant in a gene may not only be present in an ASD individual, but it may also be present in seemingly unaffected family members. Similarly, the pathogenic variant may also be observed in seemingly unaffected individuals in the general population. This phenomenon, referred to as *incomplete penetrance*, complicates the interpretation of genetic evalaution.

Alternatively, a genetic variant may result in a range of disease severity in affected individuals, a phenomenon known as *variable expressivity*. For example, ~500 kb deletions and duplications at the 16p11.2 locus are among the most heavily studied ASD-associated CNVs. However, CNVs at this locus are also responsible for a range of other neurodevelopmental and neuro-psychiatric disorders, such as schizophrenia. CNVs at the 16p11.2 locus can also be inherited from seemingly unaffected family members and have been observed in unaffected individuals in the general population. This lack of correlation between genotype and phenotype as it relates to ASD-associated genetic variation may be in part due to differences in gene-environment interactions between individuals carrying such variation.

A recent report highlights some of the challenges inherent in the genetic evaluation of ASD individuals. A putative disruptive variant in the ASD-associated *SHANK3* gene was identified in a boy with autism [52]. The variant, which was inherited from a healthy mother, was a small insertion that would be predicted to result in a frameshift and premature stop. Based on this evidence, as well as the relatively high penetrance of *SHANK3* mutations in ASD and other neuropsychiatric diseases, one could conclude that this variant in the *SHANK3* gene was pathogenically relevant in this autistic male. However, follow-up studies revealed that this variant was unlikely to be present in the majority of *SHANK3* transcripts due to alternative splicing events. Furthermore, this variant was observed in 4 out of 382 control individuals without neuropsychiatric conditions, a rate >1%. This report not only illustrates the necessity of determining the frequency of a given potentially pathogenic variant in the general population but also warns against relying too heavily on computational, or *in silico,* predictions of the effects of that variant on gene function.

5.2. Ethical considerations in the reporting of ASD genetic screening results

Informed consent, and the extent to which participants have been sufficiently informed as to the purpose of a research or clinical study, has long been an issue in the field of genetic evaluation. For example, the extent to which research findings will be released and made available is one that participants in genetic evaluation studies should be informed of beforehand. In some cases, the participants themselves may not be able to gain access to the findings of genetic evaluation. This is an issue that has only been compounded with the rise of NGS and CMA and subsequent explosion in the amount of genetic data generated by these techniques.

The sheer volume of genetic information generated by NGS and CMA not only leads to the identification of potential genetic causes of a disease of interest, but also frequently leads to the detection of other variants that are no directly related to the disease under investigation but are related to other inherited human diseases. The extent to which these incidental findings should be reported is a subject of some controversy, particularly in those situations in which genetic predisposition to an adult-onset disease is discovered in a child being evaluated for genetic causes of childhood developmental disorders. One such situation was described in a recent news feature in *Nature* in which the family of a child who had undergone genetic testing for developmental disability had to be informed that the child carried a genetic predisposition to colon cancer after extensive debate between clinical geneticists and ethics reviewers as to the extent to which such genetic information should be reported [53]. The degree to which clinical geneticists should report incidental findings in research participants has been considered by numerous authors [54-57], but as of yet there is not consensus. Many of these same ethical concerns must be considered in the reporting of genetic evaluation results in individuals with ASD.

Another consideration lies in the use of genetic evaluation to determine the recurrence risk in the siblings of children with ASD and in family planning [50]. Given a recent estimate that the recurrence rate of ASD in siblings may be as high as ~20% [58], the identification of inherited variants that potentially impart susceptibility to ASD is of critical importance both in identifying at-risk siblings that have not yet begun to manifest symptoms of ASD and in making informed decisions with regards to family planning.

5.3. Strategies for ASD genetic screening interpretation and reporting

The American College of Medicine Genetics released practice guidelines for the use of genetic screening techniques in the evaluation of individuals with ASD in 2008 [59]. In the years that have followed, additional practice guidleines and consensus statements discussing the use of CMA in the genetic evaluation of ASD cases have been published [25, 26]. With its increasing usage in the genetic evaluation of ASD cases, similar practice guidelines and consensus statements regarding NGS will likely be forthcoming, and strategies for the interpretation of NGS data in the evaluation of neurological diseases have recently been proposed [14]. In this section we highlight some of the factors to consider in the interpretation of genetic screening results in ASD cases.

5.3.1. Variant inheritance and segregation with ASD

One of the key determinants in the interpretation of ASD genetic screening results is the mechanism of variant inheritance and how closely that variant segregates with ASD. Genetic variation can either arise *de novo* or be transmitted from one or both parents. There has been considerable interest in the ASD research community in the pathogenic relevance of *de novo* variants, especially within the context of sporadic ASD cases.

As they have been subjected to less stringent evolutionary selection, *de novo* variants tend to be more deleterious than inherited variants, making them excellent candidates for sporadi-

cally-occurring disease [11]. An increased rate of *de novo* CNVs in sporadic cases compared to familial cases has been reported [21, 60], and rare *de novo* CNVs at specific genomic loci were found to associate with ASD in sporadic cases from the Simons Simplex Collection [24]. Exome sequencing studies using ASD cohorts have reported an increased rate of *de novo* gene-disrupting events (i.e. nonsense, splice-site, and frameshift mutations) in affected children compared to their unaffected siblings [10].

Whereas ASD genetic research is increasingly focused on *de novo* genetic variation, it should be remembered that the genetic basis of ASD was first established by studies demonstrating the high heritability of the disease, a fact that illustrates the continued importance of identifying inherited genetic variation in ASD cases. A number of inherited single gene variants and CNVs that segregate with disease in ASD families has been recently described [9, 61-64]; these and other findings clearly demonstrate the importance of identifying inherited variants that closely segregate with disease in affected families. It should be noted that determining the extent of variant segregation in ASD families can be complicated by the phenotypic heterogeneity that a given variant can cause from one affected family member to another. Furthermore, a disease-causing variant may exclusively segregate with disease in males, even if the variant does not reside on the X chromosome, as is the case with a *SHANK1* mutation identified in a four-generation ASD family [61]. Detailed family history and genetic evaluation of both affected and unaffected family members is essential in determining the signficance of both *de novo* and inherited variants in ASD cases.

5.3.2. Functional impact of variant

In addition to the mechanism of variant inheritance and variant segregation with disease, another important consideration in interpretation of genetic screening results lies in the functional impact of the variant. In many cases, especially with the use of high-throughput screening technologies, variant function is predicted *in silico*. In the case of single gene mutations, variation that results in disruption of gene function, such as nonsense mutations, splice-site mutations, or frameshift mutations that introduce premature stop codons, are strong genetic candidates, especially if such gene-disrupting variants are identified in a known ASD suscepibility gene or a gene associated with an ASD-linked pathway. The interpretation of missense mutations is more complicated and requires assessment of evolutionary conservation using phyloP or Genomic Evolutionary Rate Profiling (GERP) conservation scores, as well as scoring of the functional impact using Grantham or PolyPhen-2. However, as previously mentioned [52], dependency on *in silico* predictions for variant function, even in well characterized ASD-linked genes, can lead to false conclusions. As such, experimental functional assays are essential to accurately determine the impact of a given variant on gene expression or function of the encoded gene product.

5.3.3. Clinical correlations of the variant with ASD

Another consideration in the interpretation of genetic screening results in ASD cases is the degree of clinical correlation of a given variant with ASD. Hundreds of susceptibility genes and CNV loci linked with ASD have been identified and catalogued in online genetic databases

such as AutDB, DECIPHER, and others. The identification of a novel, potentially pathogenic variant in one of these known susceptibility genes or CNV loci would be strong evidence for a causal role. To a lesser extent, a novel variant in a gene in an ASD-associated pathway or a gene previously shown by gene expression studies to be differentially regulated in ASD tissue would be a strong candidate. Another factor to consider is the frequency of a variant of interest in healthy control populations; the absence or significantly reduced frequency of the variant of interest in unaffected individuals would offer strong evidence for a causal role.

5.4. Resources for ASD genetic screening interpretation

A number of online resources are available to aid clinical geneticists in the interpretation of genetic screening results in ASD individuals. Many of these resources are aimed at differentiating between rare, potentially ASD-specific variants and benign variants observed in the general population. In this section we will describe some of these resources in greater detail.

5.4.1. Genetic variation in control populations

Differentiating between potentially pathogenic and benign genetic variants in ASD cases requires knowledge of the degree of genetic variation that resides within seemingly unaffected individuals in the general population. A number of online resources, several of which are hosted by the National Center for Biotechnology Information (NCBI) [65], have been developed to allow clinical geneticists to visualize genetic variation identified in the general population. The genetic variation curated in these databases can range from single nucleotide polymorphisms to chromosomal structural variation and has proven invaluable in assessing the potential pathogenic relevance of novel genetic variants.

5.4.1.1. dbSNP (database of single nucleotide polymorphisms)

dbSNP (http://www.ncbi.nlm.nih.gov/snp) is a public domain database hosted by NCBI collecting a range of polymorphic genetic variation, including single nucleotide polymorphisms (SNPs), small-scale multi-base deletions or insertions (also called deletion insertion polymorphisms or DIPs), and retroposable element insertions and microsatellite repeat variations (also called short tandem repeats or STRs) [65].

5.4.1.2. 1,000 Genomes Project

The 1,000 Genomes Project (http://www.1000genomes.org/) is a consortium employing high-throughput NGS techniques for the purposes of characterizing over 95% of genetic variants located in genomic regions accessible to sequencing and occurring at an allelic freuqency of 1% or higher in each of five major population groups [66].

5.4.1.3. dbVar (database of genomic structural variation)

dbVar (http://www.ncbi.nlm.nih.gov/dbvar/) is a searchable online database hosted by NCBI containing genomic structural variation, defined by the database as inversions, balanced

translocations, and CNVs approximately 1 kb or larger in size, that has been observed in both case and control populations [65].

5.4.1.4. Database of Genomic Variants

The Database of Genomic Variants (http://dgvbeta.tcag.ca/dgv/app/home?ref=NCBI36/hg18) is an curated online database hosted by the Centre for Applied Genomic that contains structural variation, defined by the developers of the database as genomic alterations that involve segments of DNA that are larger than 50bp, in control individuals [67]. Users can search the database for genetic variants such as CNVs, insertions, inversions, and regions of uniparental disomy, as well as download database contents.

5.4.2. Genotype-phenotype association

The dbGaP public repository (http://www.ncbi.nlm.nih.gov/gap/) was created by the National Institutes of Health for the purposes of collecting individual-level genotype and phenotype data and associations between them [68]. The studies collected in dbGaP include genome-wide association studies, sequencing and diagnostic assays, and associations between genotype and non-clinical traits. Users can browse association results, utilize the Phenotype-Genotype Integrator (PheGenI) to search for phenotypic traits linked to GWAS data, and download data.

6. Conclusion

The development of lower-cost, high-throughput genome-wide genetic screening technologies has revolutionized the field of genetic evaluation and now provides clinical geneticists and researchers the opportunity to detect genetic variation in ASD individuals like never before. In doing so, the evidence for previously identified genetic susceptibility factors will expand, and novel ASD candidate genes and genomic loci will be identified, resulting in a better understanding of the genetic basis of ASD. However, precautions must be taken to ensure that genetic screening results are interpreted and reported properly.

Acknowledgements

Author details

Eric C. Larsen, Catherine Croft Swanwick and Sharmila Banerjee-Basu

MindSpec, Inc., U.S.A.

References

[1] Pinto D, et al. Functional impact of global rare copy number variation in autism spectrum disorders. Nature 2010; 466(7304): 368-372.

[2] O'Roak BJ, et al. Exome sequencing in sporadic autism spectrum disorders identifies severe de novo mutations. Nature Genetics 2011; 43(6): 585-589.

[3] Metzker ML. Sequencing technologies-the next generation. Nature Reviews Genetics 2010; 11(1): 31-46.

[4] O'Roak BJ, et al. Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. Nature 2012; 485(7397): 246-250.

[5] Talkowski ME, et al. Next-generation sequencing strategies enable routine detection of balanced chromosome rearrangements for clinical diagnostics and genetic research. American Journal of Human Genetics 2011; 88(4): 469-481.

[6] Talkowski ME, et al. Sequencing chromosomal abnormalities reveals neurodevelopmental loci that confer risk across diagnostic boundaries. Cell 2012; 149(3): 525-537.

[7] Sanders SJ, et al. De novo mutations revealed by whole-exome sequencing are strongly associated with autism. Nature 2012; 485(7397): 237-241.

[8] Neale BM, et al. Patterns and rates of exonic de novo mutations in autism spectrum disorders. Nature 2012; 485(7397): 242-245.

[9] Chahrour MH, et al. Whole-exome sequencing and homozygosity analysis implicate depolarization-regulated neuronal genes in autism. PLoS Genetics 2012; 8(4): e1002635. http://www.plosgenetics.org/article/info%3Adoi%2F10.1371%2Fjournal.pgen.1002635

[10] Iossifov I, et al. De novo gene disruptions in children on the autistic spectrum. Neuron 2012; 74(2): 285-299.

[11] Veltman JA and Brunner HG. De novo mutation in human genetic disease. Nature Reviews Genetics 2012; 13(8): 565-575.

[12] Cooper DN, et al. The nature of mechanisms of human gene mutation. In: Scriver CR et al. (eds.) The Metabolic and Molecular Bases of Inherited Disease. McGraw-Hill, New York; 1995. 7th edition, p259-291.

[13] Choi M, et al. Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. Proceeding of the National Academy of Sciences U.S.A. 2009; 106(45): 19096-19101.

[14] Foo J-N, et al. Whole-genome and whole-exome sequencing in neurological diseases. Nature Reviews Neurology 2012; doi: 10.1038/nrneurol.2012.148. http://www.nature.com/nrneurol/journal/vaop/ncurrent/full/nrneurol.2012.148.html

[15] Baker P, et al. Brief report: duplication of chromosome 15q11-13 in two individuals with autistic disorder. Journal of Autism and Developmental Disorders 1994; 24(4): 529-535.

[16] Bundey S, et al. Duplication of the 15q11-13 region in a patient with autism, epilepsy and ataxia. Developmental Medicine & Child Neurology 1994; 36(8): 736-742.

[17] Cook EH Jr., et al. Autism or atypical autism in maternally but not paternally derived proximal 15q duplication. American Journal of Human Genetics 1997; 60(4): 928-934.

[18] Carratalá F, et al. A patient with autistic disorder and a 20/22 chromosomal transloca-tion. Developmental Medicine & Child Neurology 1998; 40(7): 492-495.

[19] Goizet C, et al. Case with autistic syndrome and chromosome 22q13.3 deletion de-tected by FISH. American Journal of Medical Genetics 2000; 96(6): 839-844.

[20] Miller DT, et al. Consensus statement: chromosomal microarray is a first-tier clinical diagnostic test for individuals with developmental disabilities or congenital anoma-lies. American Journal of Human Genetics 2010; 86(5):749-764.

[21] Marshall CR, et al. Structural variation of chromosomes in autism spectrum disorder. American Journal of Human Genetics 2008; 82(2): 477-488.

[22] Glessner JT, et al. Autism genome-wide copy number variation reveals ubiquitin and neuronal genes. Nature 2009; 459(7246): 569-573.

[23] Shen Y, et al. Clinical genetic testing for patients with autism spectrum disorders. Pe-diatrics 2010; 125(4): e727-735.

[24] Sanders SJ, et al. Multiple recurrent de novo CNVs, including duplications of the 7q11.23 Williams syndrome region, are strongly associated with autism. Neuron 2011; 70(5): 863-885.

[25] Manning M and Hudgins L. Array-based technology and recommendations for uti-lization in medical genetics practice for detection of chromosomal abnormalities. Ge-netics in Medicine 2010; 12(11): 742-745.

[26] Miller DT, et al. Oligonucleotide microarrays for clinical diagnosis of copy number variation and zygosity status. Current Protocols in Human Genetics 2012; Chapter 8: Unit8.12.

[27] Basu SN, et al. AutDB: a gene reference resource for autism research. Nucleic Acids Research 2009; 37(Database issue): D832-836.

[28] Jamain S, et al. Mutations of the X-linked genes encoding neuroligins NLGN3 and NLGN4 are associated with autism. Nature Genetics 2003; 34(1): 27-29.

[29] Arking DE, et al. A common genetic variant in the neurexin superfamily member CNTNAP2 increases familial risk of autism. American Journal of Human Genetics 2008; 82(1): 160-164.

[30] Bakkaloglu B, et al. Molecular cytogenetic analysis and resequencing of contactin as‐ sociated protein-like 2 in autism spectrum disorders. American Journal of Human Genetics 2008; 82(1): 165-173.

[31] Li X, et al. Association analysis of CNTNAP2 polymorphisms with autism in the Chi‐ nese Han population. Psychiatric Genetics 2010; 20(3): 113-117.

[32] Strauss KA, et al. Recessive symptomatic focal epilepsy and mutant contactin-associ‐ ated protein-like 2. New England Journal of Medicine 2006; 354(13): 1370-1377.

[33] Zweier C, et al. CNTNAP2 and NRXN1 are mutated in autosomal-recessive Pitt-Hopkins-like mental retardation and determine the level of a common synaptic pro‐ tein in Drosophila. American Journal of Human Genetics 2009; 85(5): 655-666.

[34] El-Fishawy P and State MW. The genetics of autism: key issues, recent findings, and clinical implications. The Psychiatric Clinics of North America 2010; 33(1): 83-105.

[35] van der Zwaag B, et al. Gene-network analysis identifies susceptibility genes related to glycobiology in autism. PLoS One 2009; 4(5): e5324. http://www.plosone.org/arti‐ cle/info%3Adoi%2F10.1371%2Fjournal.pone.0005324

[36] Sbacchi S, et al. Functional annotation of genes overlapping copy number variants in autistic patients: focus on axon pathfinding. Current Genomics 2010; 11(2): 136-145.

[37] Gilman SR, et al. Rare de novo variants associated with autism implicate a large functional network of genes involved in formation and function of synapses. Neuron 2011; 70(5): 898-907.

[38] Anney RJ, et al. Gene-ontology enrichment analysis in two independent family-based samples highlights biologically plausible processes for autism spectrum disorders. European Journal of Human Genetics 2011; 19(10): 1082-1089.

[39] Skafidas E, et al. Predicting the diagnosis of autism spectrum disorder using gene pathway analysis. Molecular Psychiatry 2012; doi: 10.1038/mp.2012.126. http:// www.nature.com/mp/journal/vaop/ncurrent/full/mp2012126a.html

[40] Kumar A, et al. A brain region-specific predictive gene map for autism derived by profiling a reference gene set. PLoS One 2011; 6(12): e28431. http://www.plosone.org/ article/info%3Adoi%2F10.1371%2Fjournal.pone.0028431

[41] Kumar A, et al. Animal model integration to AutDB, a genetic database for autism. BMC Medical Genetics 2011; 4:15. http://www.biomedcentral.com/1755-8794/4/15

[42] Abrahams BS, et al. SFARI Gene 2.0: a community-driven knowledgebase for the Au‐ tism Spectrum Disorders (ASDs). Submitted to Nature Genetics, manuscript under review.

[43] Firth HV, et al. DECIPHER: Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources. American Journal of Human Genetics 2009; 84(4): 524-533.

[44] Xu LM, et al. AutismKB: an evidence-based knowledgebase of autism genetics. Nucleic Acids Research 2012; 40(Database issue): D1016-22.

[45] Matuszek G and Talebizadeh Z. Autism Genetic Database (AGD): a comprehensive database including autism susceptibility gene-CNVs integrated with known noncoding RNAs and fragile sites. BMC Medical Genetics 2009; 10: 102. http://www.biomed-central.com/1471-2350/10/102

[46] Nakatani J, et al. Abnormal behavior in a chromosome-engineered mouse model for human 15q11-13 duplication seen in autism. Cell 2009; 137(7): 1235-1246.

[47] Kerin T, et al. A noncoding RNA antisense to moesin at 5p14.1 in autism. Science Translational Medicine 2012; 4(128): 128ra40. http://stm.sciencemag.org/content/4/128/128ra40.short

[48] Smith CL, et al. Genomic and epigenomic instability, fragile sites, schizophrenia and autism. Current Genomics 2010; 11(6): 447-469.

[49] Levenson D. Too many children with autism miss genetics evaluations. American Journal of Medical Genetics, Part A 2012; 158A(8): vii-viii.

[50] Trottier M,,et al. Parents' perspectives on participating in genetic research in autism. Journal of Autism and Developmental Disorders 2012; http://www.springerlink.com/content/0q66420l3v518r31/?MUD=MP

[51] Gurrieri F. Working up autism: the practical role of medical genetics. American Journal of Medical Genetics. Part C, Seminars in Medical Genetics 2012; 160C(2): 104-110.

[52] Kolvezon A, et al. Analysis of a purported SHANK3 mutation in a boy with autism: clinical impact of rare variant research in neurodevelopmental disabilities. Brain Research 2011; 1380: 98-105.

[53] Hayden EC. The broken contract. Nature 2012; 486(7403): 312-314.

[54] McGuire AL, et al. Research ethics and the challenge of whole-genome sequencing. Nature Reviews Genetics 2008; 9(2): 152-156.

[55] Caulfield T, et al. Research ethics recommendations for whole-genome research: consensus statement. PLoS Biology 2008; 6(3): e73. http://www.plosbiology.org/article/info%3Adoi%2F10.1371%2Fjournal.pbio.0060073

[56] Wolf SM, et al. Managing incidental findings in human subjects research: analysis and recommendations. The Journal of Law, Medicine & Ethics 2008; 36(2): 219-48.

[57] Bredenoord AL, et al. Disclosure of individual genetic data to research participants: the debate reconsidered. Trends in Genetics 2011; 27(2): 41-47.

[58] Ozonoff S, et al. Recurrence risk for autism spectrum disorders: A baby siblings research consortium study. Pediatrics 2011; 128(3):e488-95. http://pediatrics.aappublications.org.ezproxy.nihlibrary.nih.gov/content/128/3/e488.long

[59] Schaefer GB, et al. Clinical genetics evaluation in identifying the etiology of autism spectrum disorders. Genetics in Medicine 2008; 10(4): 301-305.

[60] Sebat J, et al. Strong association of de novo copy number mutations with autism. Science 2007; 316(5823): 445-449.

[61] Sato D, et al. SHANK1 Deletions in Males with Autism Spectrum Disorder. American Journal of Human Genetics 2012; 90(5): 879-887.

[62] Duong L, et al. Mutations in NRXN1 in a family multiply affected with brain disorders: NRXN1 mutations and brain disorders. American Journal of Medical Genetics. Part B, Neuropsychiatric Genetics 2012; 159B(3): 354-358.

[63] Salyakina D, et al. Copy number variants in extended autism spectrum disorder families reveal candidates potentially involved in autism risk. PLoS One 2011; 6(10): e26049. http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0026049

[64] Shen Y, et al. Intra-family phenotypic heterogeneity of 16p11.2 deletion carriers in a three-generation Chinese family. American Journal of Medical Genetics. Part B, Neuropsychiatric Genetics 2011; 156(2): 225-232.

[65] Sayers EW, et al. Database resources of the National Center for Biotechnology Information. Nucleic Acids Research 2011; 39(Database issue): D38-51.

[66] The 1000 Genomes Project Consortium. A map of human genome variation from population-scale sequencing. Nature 467(7319): 1061-1073.

[67] Iafrate AJ, et al. Detection of large-scale variation in the human genome. Nature Genetics 2004; 36(9): 949-951.

[68] Mailman MD, et al. The NCBI dbGaP database of genotypes and phenotypes. Nature Genetics 2007; 39(10): 1181-1186.