Genet. Res., Camb. (2013), **95**, *pp.* 121–129.
doi:10.1017/S0016672313000153

High-throughput sequencing of autism spectrum disorders comes of age

MINGBANG WANG[1]*†, XIAOMEI FAN[1]†, TAO WANG[2] AND JINYU WU[2]*
[1]*BGI-tech, BGI-Shenzhen, Shenzhen 518083, China*
[2]*Wenzhou Medical College, Institute of Genomic Medicine, Wenzhou 325035, China*

(*Received 21 June 2013; revised 20 August 2013; accepted 20 August 2013*)

Summary

Autism spectrum disorders (ASDs) are lifelong neurodevelopmental disabilities that affect 1 in 88 children in the USA. Despite the high heritability, the genetic basis for a majority of the ASDs remains elusive. The considerable clinical and genetic heterogeneity pose a significant challenge technically. State-of-the-art high-throughput sequencing (HTS), which makes the analyses of any specific single/multiple genes or whole exomes feasible, has shown a promising perspective in disease gene discovery. To date, numerous genetic studies using HTS have been reported and many rare inherited or *de novo* mutations have been identified. This review will focus on the progress and prospective of genome studies of ASDs using HTS.

Introduction

Autism spectrum disorders (ASDs) are a group of lifelong and devastating neurodevelopmental disabilities that affect 1 in 88 children in the USA (Autism, Developmental Disabilities Monitoring Network Surveillance Year Principal I, Centers for Disease C, & Prevention, 2012). Multiple lines of evidence show that ASDs have a strong genetic basis. Currently, there are no effective treatments available for ASDs. Deciphering the genetic code of the ASDs would benefit the affected children through early diagnosis and intervention (Rogers *et al*., 2012), and identify new biomarkers for drug development and new therapy (Silverman *et al*., 2012). After decades of efforts, many ASDs-related syndromic genes, rare penetrant genes, rare chromosome abnormalities and structural variations have been identified, but the heritability was still missed largely (>70%) (Devlin & Scherer, 2012). The state-of-the-art high-throughput sequencing (HTS) technique already showed promise in disease-causing genes discovery not only for monogenic disorders but also for genetic diseases with complex inheritance (Rehm, 2013). This review will focus on the ASDs genome studies using HTS.

* Corresponding authors: BGI-tech, BGI-Shenzhen, Shenzhen 518083, China. E-mail: wangmingbang@genomics.cn and iamwujy@qq.cn
† These authors contributed equally to this work.

ASDs, problems and hope

The alarmingly high prevalence of ASDs poses a serious public health challenge globally and brings a high burden to the ASDs families. Multiple lines of evidence suggested a strong genetic factor in the aetiology. However, the considerable clinical and genetic heterogeneity have made the identification of genetic aetiology studies of ASDs more complex than expected. Recent studies have identified several *de novo* events, that is single nucleotide variants (SNVs), insertion and deletions (InDels), copy-number variations (CNVs) or structural variations (SVs), with high frequency among ASDs families (Veltman & Brunner, 2012). In addition, new clues also revealed some new environmental risks such as prenatal nutritional factors to ASDs (Grabrucker, 2012), including microbial infection (Heijtz *et al*., 2011) and epigenetic regulation (Shulha *et al*., 2012).

In a clinical setting, the diagnosis is mostly made based on behavioural evaluation. Recently, the application of copy number variant by chromosome microarray has become the standard of care in clinical evaluation of ASDs, and is widely used for clinical diagnosis of ASDs at an early stage (Miller *et al*., 2010). The gene panel based on syndromic gene and non-syndromic gene with high penetrance holds great prospects. Behavioural treatment such as the Early Start Denver Model (ESDM), has been reported to be effective if used early (Rogers *et al*., 2012).

A complex genetic architecture of ASDs

Four types of genetic defects have been identified in ASDs (Devlin & Scherer, 2012): (i) single genes linked to synromic ASDs (Betancur, 2011), such as *FMR1* that causes fragile X syndrome (OMIM: 309548) associated with ASDs with high frequency. These genes can explain ~10% of individuals with ASDs; (ii) rare chromosome abnormalities (Marshall *et al*., 2008) that were observed in ~5% ASDs cases; (iii) rare CNVs that contribute to ~5% cases (Malhotra & Sebat, 2012); (iv) rare penetrant genes that account for ~5% cases (Banerjee-Basu & Packer, 2010; State, 2010; Xu *et al*., 2012). As we still miss the molecular basis for the rest ~70% cases, the studies focusing on common variants have not identified any significant sites associated with ASDs (Anney *et al*., 2012). Recently, state-of-the-art HTS technology has shown promise in disease gene discovery and was applied to detect rare *de novo* or inherited variants linked to ASDs.

The progress of HTS-based ASDs studies

The recent advance in target-enrichment strategies (Mamanova *et al*., 2010) and HTS has made the detection of rare variants linked to ASDs feasible. Nowadays, one can focus on single genes, multiple genes, important pathway genes, chromosome X genes, or use whole exome sequencing (WES) and whole genome sequencing (WGS) to study genes at genome width.

Candidate gene sequencing on ASDs

For candidate gene sequencing strategy, one can focus on a single reported ASDs gene and try to find new risk variants, centre on multiple ASDs candidate genes for validation of previous results, focus on important pathway genes and try to identify new susceptible genes, or study the X-linked susceptible genes by targeting the whole X chromosome. *AFF2* is a known fragile X syndrome gene, duplications or deletions at the *AFF2* locus have also been reported in cases with moderate intellectual disability and ASDs. Mondal *et al*. (2012) sequenced the *AFF2* genomic region in 202 male ASDs probands. They found that 2·5% of males had missense mutations at highly conserved evolutionary sites and identified rare *AFF2* 3′UTR variants at conserved sites, which alter gene expression in a luciferase assay. These new identified rare variants in *AFF2* may be new ASDs susceptibility locus. A large amount of ASDs candidate genes identified by large-scale sequencing project provide a source to further study the aetiology of the ASDs. O'Roak *et al*. (2012*a*) used an ultra-low-cost molecular inversion probe to capture and sequence 44 candidate genes, which were found in the previous WES study (O'Roak *et al*., 2011, 2012*b*), in 2446 ASDs probands, the results show several new clues: (i) 27 *de novo* events in 16 genes, 59% of which are predicted to truncate proteins or disrupt splicing; (ii) recurrent disruptive mutations are observed in six genes – *CHD8*, *DYRK1A*, *GRIN2B*, *TBR1*, *PTEN* and *TBL1XR1*; (iii) these *de novo* events may contribute to 1% of sporadic ASDs; and (iv) associations between specific genes and reciprocal subphenotypes (*CHD8*-macrocephaly, *DYRK1A*-microcephaly). This new discovery may highlight the importance of a *β*-catenin/chromatin remodelling network to ASDs aetiology (Talkowski *et al*., 2012).

Many important biological pathways have been found associated with ASDs (Sakai *et al*., 2011). Study of the gene implicated in an important pathway became a good choice for studying potential risk genes or variants. The metabotropic glutamate-receptors (mGluRs) signalling pathway is one of the important pathways that has been linked to the pathophysiology of the ASDs. Several syndromic and non-syndromic genes, such as *TSC1*, *TSC2* and *SHANK3*, have been linked to ASDs. Most preclinical drugs, though just on the horizon, are based on mGluRs. To study most of the potential rare mGluRs variants linked to ASDs, Kelleher 3rd *et al*. (2012) analysed 16 genes encoding proteins in mGluRs-related pathways from pooled samples of 290 non-syndromic autism cases and 300 ethnically matched controls by using two independent next generation platforms. They (i) identified *HOMER1*, as a novel ASDs-risk gene, which encodes a scaffolding protein at the post-synaptic density-localized scaffolding protein that interacts with SHANK3 to regulate mGluRs activity; and (ii) found rare ASDs-associated coding variants predicted to have damaging effects on components of the Ras/MAPK cascade. These discoveries will benefit the mGluRs-based ASDs drug development activity.

The observed 1/4 (female/male) ratio among ASDs suggested that X-linked genes may be implicated although there is no evidence from the genetic linkage studies so far. Followed by enriching the whole chromosome X exome using a custom target region capture array, Nava *et al*. (2012) sequenced 12 unrelated families with two males affected with ASDs and identified (i) 36 possibly deleterious variants in 33 candidate genes; (ii) damaging mutations in *PHF8* and *HUWE1* that were previously implicated in intellectual disability (ID); (iii) a non sense mutation in *TMLHE* in two brothers with autism and ID; and (iv) *TMLHE* linked to ASDs by a large-scale Sanger sequencing-based screening and functional analysis.

WES or WGS studies on ASDs

WES has been widely used to target disease gene of unknown human disorder (Gilissen *et al*., 2012). As the cost for human WGS is going down very

https://doi.org/10.1017/S0016672313000153 Published online by Cambridge University Press

rapidly, one can afford to study the whole genome of a common disorder (Anney *et al*., 2012). To date, several large-scale ASDs WES studies have been published; a brief summary of these studies is shown in Table 1.

The achievements of HTS studies on ASDs

Identified amount of rare de novo point mutations linked to ASDs

De novo mutations have been found in many human genetic diseases (Veltman & Brunner, 2012); to study *de novo* mutations linked to ASDs, one can use WES of several or large-scale trios. Many rare *de novo* damaging mutations linked to ASDs have been identified to date.

O'Roak *et al*. (2011) performed the first trios-based WES on ASDs, after excluding CNVs by the CGH array, they sequenced 20 trios and (i) detected a total of 21 *de novo* variants and 11 were protein alterations, which occurred on different genes and were predicted to be damaging; (ii) suggested that the probands harboured more highly conserved and disruptive amino acid mutations than controls; and (iii) identified four potentially causative *de novo* mutations, *GRIN2B* (IVS9-2A > G), *SCN1A*(p.Pro1894Leu), *LAMC3* (p.Asp399Gly) and *FOXP1*(p.Ala339Ser fsX4), as new candidates linked to ASDs.

O'Roak *et al*. (2012*b*) selected a total of 677 individual exomes, including 189 new trios and 20 trios that were previously reported (O'Roak *et al*., 2011), for WES. They identified that (i) *de novo* point mutations were overwhelmingly paternal in origin (4:1 bias) and positively correlated with paternal age; (ii) 39% (49 of 126) of the most severe or disruptive *de novo* mutations mapped to highly interconnected b-catenin/chromatin remodelling protein network genes; (iii) two genes, *CHD8* and *NTNG1*, were found with recurrent damaging mutations in probands; and (iv) mutation screening of six candidate genes (four reported previously (O'Roak *et al*., 2011)) in 1703 ASDs probands identified additional *de novo* and protein-altering mutations in *GRIN2B*, *LAMC3* and *SCN1A*.

Sanders *et al*. (2012) performed whole-exome sequencing of 928 individuals in 238 families, including 200 discordant sibling pairs, and observed highly disruptive (nonsense and splice-site) *de novo* mutations in brain-expressed genes that are associated with ASDs and a total of 279 *de novo* coding mutations. Although there is only one single instance in the probands and none in the siblings, *SCN2A* which harboured two independent nonsense variants in probands may linked to ASDs.

Neale *et al*. (2012) sequenced the exomes of 175 trios and found that: (i) nearly half of the cases (46·3%) carry missense or nonsense *de novo* variants; (ii) the overall rate of mutation is $\sim 2 \times 10^{-8}$ per base per generation and only modestly higher than expected; (iii) the disrupted genes were found to be connected with previously reported ASDs genes via a protein–protein interaction analysis; and (iv) *CHD8* and *KATNAL2* may link to ASDs which need further validation.

Iossifov *et al*. (2012) sequenced 343 family 'quads' (the parents of a single child on the autism spectrum and its unaffected sibling), although the numbers of *de novo* missense mutations between affected and unaffected children are similar, gene-disrupting mutations (nonsense, splice site and frame shifts) are found twice as frequently, 59 to 28, in ASDs children. These mutations were mostly from the paternal line and in an age-dependent manner. Many of these disrupted genes are associated with the fragile X protein, FMRP.

Identified multiples of rare inherited point mutations, may lead to cure of ASDs

To study rare inherited mutations linked to ASDs, WES of several probands or large-scale case–control samples can be applied.

Chahrour *et al*. (2012) sequenced the whole exomes of 16 probands from 16 non-consanguineous families that showed evidence of distant shared ancestry. By using the homozygosity analysis and multidimensional strategy for filtering whole-exome sequence data to find candidate recessive mutations, four candidate genes (e.g. *UBE3B*, *CLTCL1*, *NCKAP5L* and *ZNF18*) were identified, and the expression of these genes were highly responsive to neuronal activity.

Cheng *et al*. (2012) performed whole-exome sequence and extensive bioinformatic analysis of a cohort of 20 ASDs patients, identified novel mutations in seven genes that are implicated in synaptic function and neurodevelopment. After sequencing an additional 47 ASDs samples, they identified three different missense mutations in *ANK3* in four unrelated ASDs patients, one of which, c.4705T > G (p.S1569A), is a *de novo* mutation. There is shared molecular pathophysiology between the ASDs and other neuropsychiatric disorders considering that *ANK3* was previously reported to be linked to Schizophrenia and Bipolar Disorder (Schizophrenia Psychiatric Genome-Wide Association Study C, 2011). The *ANK3* was replicated by Lingling *et al*. (2013) through a WGS study of two probands from a large pedigree (including two parents and eight children). They (i) identified a list of 59 candidate variants that may increase susceptibility to autism; (ii) manual examination of this list identified *ANK3* as the most likely candidate gene; and (iii) identified 33 prioritized

https://doi.org/10.1017/S0016672313000153 Published online by Cambridge University Press

Table 1. *Summary of current WES and WGS studies on ASDs*

Epub date	Journal	Suppose	Study design	Main achievements	Reference
2011/5/17	*Nat Genet*	*De novo* variants may contribute to ASDs	WES of 20 trios, call *de novo* mutations, validation and functional characterization of FOXP1	Trio-based WES worked; Identified interest of *de novo* mutations in *FOXP1*, *GRIN2B*, *SCN1A* and *LAMC3*	O'Roak *et al.* (2011)
2012/4/13	*Nature*	*De novo* variants may contribute to ASDs	WES of 175 trios, *de novo* mutations calling, validation by Sanger sequencing, pathway and protein–protein interaction analysis	Nearly half of the cases (46·3%) carry a missense or nonsense *de novo* variants; *CHD8* and *KATNAL2* are risk factors of ASDs	Neale *et al.* (2012)
2012/4/13	*Nature*	*De novo* variants may contribute to ASDs	WES of 677 individual from 209 families (189 new trios, 20 previous trios (O'Roak *et al.*, 2011), and 50 unaffected siblings), *de novo* mutations calling, Sanger confirming and mutation screening of six candidate genes in 1703 ASDs	Observe recurrent mutation in *CHD8* and *NTNG1*, and *de novo*, protein-altering mutations in *GRIN2B, LAMC3* and *SCN1A*	O'Roak *et al.* (2012*b*)
2012/4/13	*Nature*	*De novo* variants may contribute to ASDs	WES of 928 individuals from 238 families (including 200 phenotypically discordant sibling pairs), *de novo* mutations analysis and validation	Two independent nonsense variants disrupt *SCN2A*	Sanders *et al.* (2012)
2012/4/19	*PLoS Genet*	Recessive inherited variants may contribute to ASDs	WES of 16 probands, homozygous variants and recessive inheritance analysis, validate interest variants by Sequenom genotyping and neuronal activity analysis	Identified four potential candidate genes, *UBE3B*, *CLTCL1*, *NCKAP5L* and *ZNF18*	Chahrour *et al.* (2012)
2012/4/24	*Cell*	Rare chromosomal abnormalities contribute to ASDs and NDDs	WGS of 38 patients with autism or related NDDs by an optimized jumping library, CNV calling	33 loci were found	Talkowski *et al.* (2012)
2012/5/1	*Neuron*	*De novo* variants may contribute to ASDs	WES of 343 families, *de novo* mutations analysis and validation	*FMRP*-associated genes are under greater purifying selection than the remainder of genes	Iossifov *et al.* (2012)
2012/8/7	*Hum Mutat*	Recessive inherited variants may contribute to ASDs	WES of 20 ASDs patients, filter normal variants and validation risk variants in additional 47 ASDs samples by Sanger sequencing	Identified three missense mutations in *ANK3*, one is *de novo* mutation	Cheng *et al.* (2012)
2012/8/24	*Nature*	Father's age contributes to ASDs	78 Icelandic parent–offspring trios	73 are exonic *de novo* mutations, *CUL3* and *EPHB2* may link to ASDs, mutation rate is dominated by the age of the father at conception	Kong *et al.* (2012)
2012/9/8	*Science*	Recessive inherited variants may contribute to ASDs	WES of six probands with autism, epilepsy and ID from three consanguineous families, homozygous variants and recessive inheritance analysis, functional studies using protein, expression, cell culture, knockout mice etc.	Autism with ID and epilepsy caused by *BCKDK* and represents a potentially treatable syndrome	Novarino *et al.* (2012)
2012/12/25	*Cell*	*De novo* variants may contribute to ASDs	WGS of 10 MZ twins with ASDs and their parents, *de novo* mutation analysis by ForestDNM (machine learning-based Algorithm to call *de novo* mutations), Sanger sequencing validation and annotation	Mutation rates varied widely throughout the genome (by 100-fold); dense clusters of mutations within individual genomes were attributable to compound mutation or gene conversion; hypermutability was one of the characteristics of genes involved in ASDs and other diseases	Michaelson *et al.* (2012)
2013/1/29	*Neuron*	Recessive-inherited variants may contribute to ASDs	WES of 933 cases and 869 controls, homozygous or compound heterozygous loss-of-function (LoF) variants analysis, and validation in 563 probands and 4605 controls	Identified rare inherited autosomal and X chromosome complete gene knockouts as risk factors of ASDs	Lim *et al.* (2013)

2013/1/29	*Neuron*	Recessive inherited variants may contribute to ASDs	WES of 3 consanguineous and/or multiplex ASDs families, TRS of additional 70 genes with neurocognitive effect in 163 consanguineous and/or multiplex families, and validation in published 612 non-consanguineous families (Iossifov *et al.*, 2012; O'Roak *et al.*, 2012*b*; Sanders *et al.*, 2012)	Six genes (*AMT*, *PEX7*, *SYNE1*, *VPS13B*, *PAH* and *POMGNT1*) with inherited biallelic mutations were found to have been associated with ASDs	Yu *et al.* (2013)
2013/4/18	*PLoS Genet*	Recessive inherited variants may contribute to ASDs	WES of 1039 cases and 870 controls, from two centres, gene-based tests and both meta-(data analysed and then be combined) and mega-analysis (data be combined and then analysed)	Mega-analysis is better than meta-analysis; found no new risk genes; gene-based tests will require much larger samples	Liu *et al.* (2013)
2013/4/20	*Mol Autism*	Recessive inherited variants may contribute to ASDs	WGS of two probands, recessive model of inheritance analysis, Sanger sequencing validation	*ANK3* as the most likely candidate gene	Lingling *et al.* (2013)
2013/7/16	*Am J Hum Genet*	Recessive inherited or *de novo* variants may contribute to ASDs	WGS of 32 trios, *de novo* mutations and inheritance analysis, Sanger sequencing validation and families analysis	Identified *de novo* mutations in six of 32 (19%) families and X-linked or autosomal inherited alterations in ten of 32 (31%) families; found four unrecognized (*KAL*, *CAPRIN1*, *VIP*, *KCNQ2*), nine known (*AFF2*, *ARHGEF6*, *DMD*, *CACNA1C*, *CHD7*, *EHMT1*, *SATB2*, *SCN2A*, *NRXN1*) and eight candidate genes (*BCORL1*, *WWC3*, *ZC3H12B*, *ARID5A*, *DNMT3A*, *KIAA0284*, *USP54*, and *MICALCL*) linked to ASDs	Jiang *et al.* (2013)

non-coding variants such as those near *SMG6* and *COQ5*, based on evolutionary constraint and experimental evidence from ENCODE.

Yu *et al.* (2013) firstly apply WES to a cohort of three consanguineous and/or multiplex ASDs families which shared ancestry between the parents, typically as cousins; then they used TRS to screen for mutations in additional 70 genes with neurocognitive effect in a total of 163 consanguineous and/or multiplex families. Lastly, they used 612 published non-consanguineous families data for validation, finally six genes (e.g., *AMT*, *PEX7*, *SYNE1*, *VPS13B*, *PAH* and *POMGNT1*) with inherited biallelic mutations were found to have been associated with ASDs.

The identified rare variants may lead to better cure of ASDs. To develop a better cure for ASDs, Novarino *et al.* (2012) performed whole-exome sequencing from two consanguineous families, one of Turkish descent and a second of Egyptian ancestry, and identified inactivating mutations in the gene Branched Chain Ketoacid Dehydrogenase Kinase (*BCKDK*) in consanguineous families with autism, epilepsy and ID. The patients with homozygous *BCKDK* mutations display reductions in *BCKDK* mRNA and protein, and *BCKDK* knockout mice show abnormal brain amino acid profiles and neurobehavioural deficit. Nevertheless, the deficit is reversible using dietary supplementation. This may represent a potentially treatable syndrome in ASDs.

De novo mutations varied widely in ASDs and were associated with father's age

Kong *et al.* (2012) conducted a study of genome-wide mutation rates by WGS of 78 Icelandic parent–offspring trios at high coverage (30X average coverage). They show that (i) with an average father's age of 29·7, the average *de novo* mutation rate is $1{\cdot}2\times10^{-8}$ per nucleotide per generation; (ii) 73 are exonic *de novo* mutations, among which *CUL3* and *EPHB2* may link to ASDs and (iii) the diversity in mutation rate of single nucleotide polymorphisms is dominated by the age of the father at conception of the child and father's age.

Michaelson *et al.* (2012) firstly developed a machine-learning based *de novo* mutation calling pipeline, ForestDNM, by which they analysed *de novo* point mutations in ten identical twin pairs and their parents. Results showed that: (i) mutation rates varied widely throughout the genome (by 100-fold) and could be explained by the intrinsic characteristics of the DNA sequence and chromatin structure; (ii) dense clusters of mutations within individual genomes were attributable to compound mutation or gene conversion; and (iii) hypermutability was one of the characteristics of genes involved in ASDs and other diseases.

https://doi.org/10.1017/S0016672313000153 Published online by Cambridge University Press

Failed to identify common variants linked to ASDs

Although Genome Wide Association Studies (GWAS) have not made a significant contribution to ASDs gene discoveries, large-scale WES-based case-control study shows promise on common diseases as they can focus on sites with MAF < 5%. Liu *et al.* (2013) evaluated the association of rare variants and ASDs in 1039 cases and 870 controls with similar ancestry, about half of the samples were sequenced on the Solid platform (Baylor: 505 cases, 491 controls) and the remainder were sequenced on the Illumina platform (Broad: 534 cases, 379 controls). Gene-based association analyses were conducted but no gene showed exome-wide significant association. This may suggest that rare risk variants are scattered across these many genes, and larger samples would be required.

WGS holds promise to identify rare chromosome abnormalities

To study rare chromosome abnormalities associated with ASDs, especially balanced chromosomal abnormalities (BCAs), which represent a relatively untapped reservoir of single gene disruptions in neurodevelopmental disorders (NDDs), a cost-effective pipeline was developed to call genomic rearrangements and structural variations at base pair resolution (Talkowski *et al.*, 2011). Talkowski *et al.* (2012) sequenced BCAs in 38 patients with autism or related NDDs by using an optimized jumping library protocol (21 subjects), a targeted DNA capture protocol (10 subjects), a standard paired-end (PE) protocol (two subjects), and an Illumina mate-pair (MP) protocol (five subjects). As a consequence, disruptions of 33 loci were found in four general categories: (1) genes previously associated with abnormal neurodevelopment (e.g. *AUTS2*, *FOXP1* and *CDKL5*); (2) single-gene contributed to microdeletion syndromes (*MBD5*, *SATB2*, *EHMT1* and *SNURF-SNRPN*); (3) novel risk loci (e.g. *CHD8*, *KIRREL3* and *ZNF507*); and (4) genes associated with later-onset psychiatric disorders (e.g. *TCF4*, *ZNF804A*, *PDE10A*, *GRIN2B* and *ANK3*). By evaluation of CNVs in independent subjects, significantly increased burdens of CNVs from these 33 loci were found. By performing gene-set enrichment analysis using published GWAS data and network analysis, significant enrichments of polygenic risk alleles in these 33 loci were also found.

WGS shows better performance to identify rare variants linked to ASDs

Jiang *et al.* (2013) used WGS to study 32 ASDs trios; after *de novo* mutations and inheritance analysis, they (i) identified that deleterious *de novo* mutations in six of 32 (19%) families and X-linked or autosomal inherited alterations in 10 of 32 (31%) families, may be due to the comprehensive and uniform coverage afforded by WGS; (ii) deleterious variants were found in four unrecognized (e.g. *KAL*, *CAPRIN1*, *VIP* and *KCNQ2*), nine known ASDs genes (e.g. *AFF2*, *ARHGEF6*, *DMD*, *CACNA1C*, *CHD7*, *EHMT1*, *SATB2*, *SCN2A* and *NRXN1*) and eight candidate ASDs risk genes (e.g. *BCORL1*, *WWC3*, *ZC3H12B*, *ARID5A*, *DNMT3A*, *KIAA0284*, *USP54* and *MICALCL*).

Future direction

To study the epigenetic risk factor linked to ASDs

MZ twins discordant for ASDs provide an ideal model to study epigenetics risk factors of ASDs since the twins share most of their genetic background. Wong *et al.* (2013) performed a genome-wide analysis of DNA methylation in a sample of 50 MZ twin pairs (100 individuals), included twins discordant and concordant for ASDs, ASDs-associated traits and no autistic phenotype. Within-twin and between-group analyses identified numerous differentially methylated regions associated with ASDs.

To study the gene expression of ASDs genes

Study of the gene expression of ASDs holds promise in deciphering the DNA mutation effect of ASDs. As brain tissue is not available from most samples, study of the gene expression of peripheral blood mononuclear cell (PBMC) provides an easy way. Luo *et al.* (2012) interrogated gene expression in lymphoblasts from 244 families with discordant siblings by microarray.

To study the gut microbes linked to ASDs

Gut microbiomes are a new frontier in autism research (Mulle *et al.*, 2013); although no direct evidence exists, there are clues that gut microbiota modulate brain development and behaviour (Heijtz *et al.*, 2011). Ming *et al.* (2012) performed liquid- and gas-chromatography-based mass spectrometry to study metabolomics in urinary specimens from 48 children with ASDs and 53 age-matched controls. They detected abnormal amino acid metabolism, increased oxidative stress, and altered gut microbiomes in ASDs.

Prospective

The problem of ASDs has been exposed to the spotlight; several large-scale WES have led to the identification of multiple ASDs candidate genes. The WGS

https://doi.org/10.1017/S0016672313000153 Published online by Cambridge University Press

showed better detection rate not only on coding and non-coding variants (Jiang *et al*., 2013), but also on CNVs and SVs. The state-of-the-art HTS will enable us to decode the missed heritability for ASDs. The challenge is how to analyse the *de novo* or inherited event of rare variations and understand their functions.

Recently, several large-scale WGS projects have begun, one of which is called the Autism Genome 10 K project (http://www.autismgenome10k.org/) initiated by BGI, the world's largest genomic organization, and Autism Speaks, the world's largest autism science and advocacy organization. The project aims to: (i) sequence the genomes from 10 000 individuals belonging to 2000 ASDs families from the Autism Speaks Autism Genetic Resource Exchange (AGRE) and 1000 Chinese ASDs families; (ii) analyse *de novo* events or inherited SNVs, InDels, CNVs, SVs etc.; (iii) identify ASDs genes or variants; and (iv) screen targets for drug development and therapy. The pilot results are published (Jiang *et al*., 2013) and the planned transomics study of ASDs, which include transcriptome, epigenome, proteome, metabolome and metagenome study, will focus on the ASDs families which lack explanation by the genetic studies. The project may bring together scientists, ASDs families, funding agencies from governments, non-government organizations which provide autism intervention services and drug companies from all over the world.

We thank Dr Yong-hui Jiang and Dr Zhongsheng Sun's suggestions for preparing and revising this manuscript.

References

Anney, R., Klei, L., Pinto, D., Almeida, J., Bacchelli, E., Baird, G., Bolshakova, N., Bolte, S., Bolton, P. F., Bourgeron, T., Brennan, S., Brian, J., Casey, J. *et al*. (2012). Individual common variants exert weak effects on the risk for autism spectrum disorderspi. *Human Molecular Genetics* **21**, 4781–4792.

Autism, Developmental Disabilities Monitoring Network Surveillance Year Principal I, Centers for Disease C & Prevention (2012). Prevalence of autism spectrum disorders – Autism and Developmental Disabilities Monitoring Network, 14 sites, United States, 2008. *Morbidity and Mortality Weekly Report. Surveillance Summaries* **61**, 1–19.

Banerjee-Basu, S. & Packer, A. (2010). SFARI Gene: an evolving database for the autism research community. *Disease Models & Mechanisms* **3**, 133–135.

Betancur, C. (2011). Etiological heterogeneity in autism spectrum disorders: more than 100 genetic and genomic disorders and still counting. *Brain Research* **1380**, 42–77.

Cheng, B. *et al*. (2012). Mutations of ANK3 identified by exome sequencing are associated with autism susceptibility. *Human Mutation* **33**, 1635–1638.

Chahrour, M. H., Yu, T. W., Lim, E. T., Ataman, B., Coulter, M. E., Hill, R. S., Stevens, C. R., Schubert, C. R., Collaboration, A. A. S., Greenberg, M. E., Gabriel, S. B. & Walsh, C. A. (2012). Whole-exome sequencing and homozygosity analysis implicate depolarization-regulated neuronal genes in autism. *PLoS Genetics* **8**, e1002635.

Devlin, B. & Scherer, S. W. (2012). Genetic architecture in autism spectrum disorder. *Current Opinion in Genetics & Development* **22**, 229–237.

Gilissen, C., Hoischen, A., Brunner, H.G. & Veltman, J.A. (2012). Disease gene identification strategies for exome sequencing. *European Journal of Human Genetics* **20**, 490–497.

Grabrucker, A. M. (2012). Environmental factors in autism. *Frontiers in Psychiatry* **3**, 118.

Heijtz, R. D., Wang, S., Anuar, F., Qian, Y., Bjorkholm, B., Samuelsson, A., Hibberd, M. L., Forssberg, H. & Pettersson, S. (2011). Normal gut microbiota modulates brain development and behavior. *Proceedings of the National Academy of Sciences USA* **108**, 3047–3052.

Iossifov, I., Ronemus, M., Levy, D., Wang, Z., Hakker, I., Rosenbaum, J., Yamrom, B., Lee, Y. H., Narzisi, G., Leotta, A., Kendall, J., Grabowska, E. *et al*. (2012). *De novo* gene disruptions in children on the autistic spectrum. *Neuron* **74**, 285–299.

Jiang, Y. H., Yuen, R. K., Jin, X., Wang, M., Chen, N., Wu, X., Ju, J., Mei, J., Shi, Y., He, M., Wang, G., Liang, J., Wang, Z. *et al*. (2013). Detection of clinically relevant genetic variants in autism spectrum disorder by whole-genome sequencing. *American Journal of Human Genetics* **93**, 249–263.

Kelleher, R. J. 3rd, Geigenmuller, U., Hovhannisyan, H., Trautman, E., Pinard, R., Rathmell, B., Carpenter, R. & Margulies, D. (2012). High- throughput sequencing of mGluR signaling pathway genes reveals enrichment of rare variants in autism. *PLoS One* **7**, e35003.

Kong, A., Frigge, M. L., Masson, G., Besenbacher, S., Sulem, P., Magnusson, G., Gudjonsson, S. A., Sigurdsson, A., Jonasdottir, A., Jonasdottir, A., Wong, W. S., Sigurdsson, G., Walters, G. B., Steinberg, S., Helgason, H., Thorleifsson, G., Gudbjartsson, D. F., Helgason, A., Magnusson, O. T., Thorsteinsdottir, U. & Stefansson, K. (2012). Rate of *de novo* mutations and the importance of father's age to disease risk. *Nature* **488**, 471–475.

Lim, E. T., Raychaudhuri, S., Sanders, S. J., Stevens, C., Sabo, A., Macarthur, D. G., Neale, B. M., Kirby, A., Ruderfer, D. M., Fromer, M., Lek, M., Liu, L. *et al*. (2013). Rare complete knockouts in humans: population distribution and significant role in autism spectrum disorders. *Neuron* **77**, 235–242.

Lingling, S. *et al*. (2013). Whole-genome sequencing in an autism multiplex family. *Molecular Autism* **4**, 8.

Liu, L., Sabo, A., Neale, B. M., Nagaswamy, U., Stevens, C., Lim, E., Bodea, C. A., Muzny, D., Reid, J. G., Banks, E., Coon, H., Depristo, M. *et al*. (2013). Analysis of rare, exonic variation amongst subjects with autism spectrum disorders and population controls. *PLoS Genetics* **9**, e1003443.

Luo, R., Sanders, S. J., Tian, Y., Voineagu, I., Huang, N., Chu, S. H., Klei, L., Cai, C., Ou, J., Lowe, J. K., Hurles, M. E., Devlin, B., State, M. W. & Geschwind, D. H. (2012). Genome-wide transcriptome profiling reveals the functional impact of rare de novo and recurrent CNVs in autism spectrum disorders. *American Journal of Human Genetics* **91**, 38–55.

Malhotra, D. & Sebat, J. (2012). CNVs: harbingers of a rare variant revolution in psychiatric genetics. *Cell* **148**, 1223–1241.

Mamanova, L., Coffey, A. J., Scott, C. E., Kozarewa, I., Turner, E. H., Kumar, A., Howard, E., Shendure, J. & Turner, D. J. (2010). Target-enrichment strategies for next-generation sequencing. *Nature Methods* **7**, 111–118.

Marshall, C. R., Noor, A., Vincent, J. B., Lionel, A. C., Feuk, L., Skaug, J., Shago, M., Moessner, R., Pinto, D., Ren, Y., Thiruvahindrapduram, B., Fiebig, A. *et al*. (2008). Structural variation of chromosomes in autism spectrum disorder. *American Journal of Human Genetics* **82**, 477–488.

Michaelson, J. J., Shi, Y., Gujral, M., Zheng, H., Malhotra, D., Jin, X., Jian, M., Liu, G., Greer, D., Bhandari, A., Wu, W. & Corominas, R. (2012). Whole-genome sequencing in autism identifies hot spots for de novo germline mutation. *Cell* **151**, 1431–1442.

Miller, D. T., Adam, M. P., Aradhya, S., Biesecker, L. G., Brothman, A. R., Carter, N. P., Church, D. M., Crolla, J. A., Eichler, E. E., Epstein, C. J., Faucett, W. A., Feuk, L. *et al*. (2010). Consensus statement: chromosomal microarray is a first–tier clinical diagnostic test for individuals with developmental disabilities or congenital anomalies. *American Journal of Human Genetics* **86**, 749–764.

Ming, X., Stein, T. P., Barnes, V., Rhodes, N. & Guo, L. (2012). Metabolic perturbance in autism spectrum disorders: a metabolomics study. *Journal of Proteome Research* **11**, 5856–5862.

Mondal, K., Ramachandran, D., Patel, V. C., Hagen, K. R., Bose, P., Cutler, D. J. & Zwick, M. E. (2012). Excess variants in AFF2 detected by massively parallel sequencing of males with autism spectrum disorder. *Human Molecular Genetics* **21**, 4356–4364.

Mulle, J. G., Sharp, W. G. & Cubells, J. F. (2013). The gut microbiome: a new frontier in autism research. *Current Psychiatry Report* **15**, 337.

Nava, C., Lamari, F., Heron, D., Mignot, C., Rastetter, A., Keren, B., Cohen, D., Faudet, A., Bouteiller, D., Gilleron, M., Jacquette, A., Whalen, S. *et al*. (2012). Analysis of the chromosome X exome in patients with autism spectrum disorders identified novel candidate genes, including TMLHE. *Translational Psychiatry* **2**, e179.

Neale, B. M., Kou, Y., Liu, L., Ma'ayan, A., Samocha, K. E., Sabo, A., Lin, C. F., Stevens, C., Wang, L. S., Makarov, V., Polak, P., Yoon, S. *et al*. (2012). Patterns and rates of exonic *de novo* mutations in autism spectrum disorders. *Nature* **485**, 242–245.

Novarino, G., El-Fishawy, P., Kayserili, H., Meguid, N. A., Scott, E. M., Schroth, J., Silhavy, J. L., Kara, M., Khalil, R. O., Ben-Omran, T., Ercan-Sencicek, A. G., Hashish, A. F., Sanders, S. J., Gupta, A. R., Hashem, H. S., Matern, D., Gabriel, S., Sweetman, L., Rahimi, Y., Harris, R. A., State, M. W. & Gleeson, J. G. (2012). Mutations in BCKD-kinase lead to a potentially treatable form of autism with epilepsy. *Science* **338**, 394–397.

O'Roak, B. J., Deriziotis, P., Lee, C., Vives, L., Schwartz, J. J., Girirajan, S., Karakoc, E., Mackenzie, A. P., Ng, S. B., Baker, C., Rieder, M. J., Nickerson, D. A., Bernier, R., Fisher, S. E., Shendure, J. & Eichler, E. E. (2011). Exome sequencing in sporadic autism spectrum disorders identifies severe *de novo* mutations. *Nature Genetics* **43**, 585–589.

O'Roak, B. J., Vives, L., Fu, W., Egertson, J. D., Stanaway, I. B., Phelps, I. G., Carvill, G., Kumar, A., Lee, C., Ankenman, K., Munson, J., Hiatt, J. B., Turner, E. H., Levy, R., O'Day, D. R., Krumm, N., Coe, B. P., Martin, B. K., Borenstein, E., Nickerson, D. A., Mefford, H. C., Doherty, D., Akey, J. M., Bernier, R., Eichler, E. E. & Shendure, J. (2012*a*). Multiplex targeted sequencing identifies recurrently mutated genes in autism spectrum disorders. *Science* **338**, 1619–1622.

O'Roak, B. J., Vives, L., Girirajan, S., Karakoc, E., Krumm, N., Coe, B. P., Levy, R., Ko, A., Lee, C., Smith, J. D., Turner, E. H., Stanaway, I. B., Vernot, B., Malig, M., Baker, C., Reilly, B., Akey, J. M., Borenstein, E., Rieder, M. J., Nickerson, D. A., Bernier, R., Shendure, J. & Eichler, E. E. (2012*b*). Sporadic autism exomes reveal a highly interconnected protein network of *de novo* mutations. *Nature* **485**, 246–250.

Rehm, H. L. (2013). Disease-targeted sequencing: a cornerstone in the clinic. *Nature Review Genetics* **14**, 295–300.

Rogers, S. J., Estes, A., Lord, C., Vismara, L., Winter, J., Fitzpatrick, A., Guo, M. & Dawson, G. (2012). Effects of a brief Early Start Denver model (ESDM)-based parent intervention on toddlers at risk for autism spectrum disorders: a randomized controlled trial. *Journal of the American Academy of Child and Adolescent Psychiatry* **51**, 1052–1065.

Sakai, Y., Shaw, C. A., Dawson, B. C., Dugas, D. V., Al-Mohtaseb, Z., Hill, D. E. & Zoghbi, H. Y. (2011). Protein interactome reveals converging molecular pathways among autism disorders. *Science Translational Medicine* **3**, 86–49.

Sanders, S. J., Murtha, M. T., Gupta, A. R., Murdoch, J. D., Raubeson, M. J., Willsey, A. J., Ercan-Sencicek, A. G., DiLullo, N. M., Parikshak, N. N., Stein, J. L., Walker, M. F., Ober, G. T., Teran, N. A. *et al*. (2012). *De novo* mutations revealed by whole-exome sequencing are strongly associated with autism. *Nature* **485**, 237–241.

Schizophrenia Psychiatric Genome–Wide Association Study C (2011). Genome-wide association study identifies five new schizophrenia loci. *Nature Genetics* **43**, 969–976.

Shulha, H. P., Cheung, I., Whittle, C., Wang, J., Virgil, D., Lin, C. L., Guo, Y., Lessard, A., Akbarian, S. & Weng, Z. (2012). Epigenetic signatures of autism: trimethylated H3K4 landscapes in prefrontal neurons. *Archives of General Psychiatry* **69**, 314–324.

Silverman, J. L., Smith, D. G., Rizzo, S. J., Karras, M. N., Turner, S. M., Tolu, S. S., Bryce, D. K., Smith, D. L., Fonseca, K., Ring, R. H. & Crawley, J. N. (2012). Negative allosteric modulation of the mGluR5 receptor reduces repetitive behaviors and rescues social deficits in mouse models of autism. *Science Translational Medicine* **4**, 131–151.

State, M. W. (2010). The genetics of child psychiatric disorders: focus on autism and Tourette syndrome. *Neuron* **68**, 254–269.

Talkowski, M. E., Ernst, C., Heilbut, A., Chiang, C., Hanscom, C., Lindgren, A., Kirby, A., Liu, S., Muddukrishna, B., Ohsumi, T. K., Shen, Y., Borowsky, M., Daly, M. J., Morton, C. C. & Gusella, J. F. (2011). Next-generation sequencing strategies enable routine detection of balanced chromosome rearrangements for clinical diagnostics and genetic research. *American Journal of Human Genetics* **88**, 469–481.

Talkowski, M. E., Rosenfeld, J. A., Blumenthal, I., Pillalamarri, V., Chiang, C., Heilbut, A., Ernst, C., Hanscom, C., Rossin, E., Lindgren, A. M., Pereira, S. & Ruderfer, D. (2012). Sequencing chromosomal

abnormalities reveals neurodevelopmental loci that confer risk across diagnostic boundaries. *Cell* **149**, 525–537.

Veltman, J. A. & Brunner, H. G. (2012). *De novo* mutations in human genetic disease. *Nature Review Genetics* **13**, 565–575.

Wong, C. C., Meaburn, E. L., Ronald, A., Price, T. S., Jeffries, A. R., Schalkwyk, L. C., Plomin, R. & Mill, J. (2013). Methylomic analysis of monozygotic twins discordant for autism spectrum disorder and related behavioural traits. *Molecular Psychiatry* doi: 10.1038/mp.2013.41.

Xu, L. M., Li, J. R., Huang, Y., Zhao, M., Tang, X. & Wei, L. (2012). Autism KB: an evidence-based knowledgebase of autism genetics. *Nucleic Acids Research* **40**, D1016–D1022.

Yu, T. W., Chahrour, M. H., Coulter, M. E., Jiralerspong, S., Okamura-Ikeda, K., Ataman, B., Schmitz-Abe, K., Harmin, D. A., Adli, M., Malik, A. N., D'Gama, A. M., Lim, E. T. *et al*. (2013). Using whole-exome sequencing to identify inherited causes of autism. *Neuron* **77**, 259–273.