

Recent Advances in Autism Spectrum Disorders: Applications of Whole Exome Sequencing Technology

Elif Funda Sener^{1,2} ✉, Halit Canatan¹, and Yusuf Ozkul³

¹Department of Medical Biology, Erciyes University Medical Faculty, Kayseri, Turkey

²Erciyes University Genome and Stem Cell Center, Kayseri, Turkey

³Department of Medical Genetics, Erciyes University Medical Faculty, Kayseri, Turkey

Autism spectrum disorders (ASD) is characterized by three core symptoms with impaired reciprocal social interaction and communication, a pattern of repetitive behavior and/or restricted interests in early childhood. The prevalence is higher in male children than in female children. As a complex neurodevelopmental disorder, the phenotype and severity of autism are extremely heterogeneous with differences from one patient to another. Genetics has a key role in the etiology of autism. Environmental factors are also interacting with the genetic profile and cause abnormal changes in neuronal development, brain growth, and functional connectivity. The term of exome represents less than 1% of the human genome, but contains 85% of known disease-causing variants. Whole-exome sequencing (WES) is an application of the next generation sequencing technology to determine the variations of all coding regions, or exons of known genes. For this reason, WES has been extensively used for clinical studies in the recent years. WES has achieved great success in the past years for identifying Mendelian disease genes. This review evaluates the potential of current findings in ASD for application in next generation sequencing technology, particularly WES. WES and whole-genome sequencing (WGS) approaches may lead to the discovery of underlying genetic factors for ASD and may thereby identify novel therapeutic targets for this disorder. **Psychiatry Investig 2016;13(3):255-264**

Key Words Autism, Autism spectrum disorders, Whole exome sequencing, Next generation sequencing, Genetics.

INTRODUCTION

Autism spectrum disorders (ASD) is characterized by impairments in reciprocal social interaction, communication and by restricted/repetitive behaviors in early childhood.^{1,2} ASD affects about 1 in 80–110 individuals, with onset before the age of three years.^{1,3,4} In addition to specific clinical symptoms, approximately 31% ASD patients' also present intellectual disability (ID) and 20–25% has seizures.⁵⁻⁷ There is no definitive pharmacotherapy for the treatment of core symptoms of ASD. Therapies commonly used to reduce behavioral, educational and immunological symptoms. Antipsychotic drugs only target the secondary symptoms of irritability, aggression, anxiety, depression, and self-injurious behaviors.^{2,8} Until now, family-

based and twin studies indicated significant genetic basis for ASD susceptibility.^{1,3} Etiopathogenesis of ASD remains largely unknown but now, it is well recognized that ASD as a complex disorder, involve the complex interaction of several genes and environmental risk factors.⁹⁻¹¹ The genetic architecture of ASD comprises a diversity of rare single nucleotide variants, copy number variations (CNVs), chromosomal abnormalities, and common polymorphic variations.¹² The recent advances in ASD genetics are summarized in Table 1. Additionally epigenetic dysregulation might contribute to significant proportion of ASD cases.^{2,11,13}

Advances in genomic technology, including array-based genome-wide association studies (GWAS) of single nucleotide polymorphisms (SNPs), CNV studies and whole exome-sequencing (WES) technologies, have been key to this paradigm shift in psychiatric research.^{12,14-16} Like other complex diseases such as hypertension and diabetes, psychiatric disorders are not inherited in a Mendelian fashion but have a complex genetic architecture involving a spectrum of mutations from small DNA sequence variations to large chromosomal rearrangements and epigenetic regulation.^{17,18} According to latest data, rare truncating heterozygous variants have a predomi-

Received: April 15, 2015 Revised: September 22, 2015

Accepted: October 2, 2015 Available online: February 17, 2016

✉ Correspondence: Elif Funda Sener, PhD

Department of Medical Biology, Erciyes University Medical Faculty, Kayseri 38039, Turkey

Tel: +90352 2076666-13600, Fax: +90352 4379316, E-mail: efefunda@yahoo.com

© This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Table 1. Recent key advances in the genetic studies of ASD

Year/period	Methods	Related events
1943	-	Identification of autism by Leo Kanner
1970–	Twin studies	Genetic contribution of ASD (60–90% heritability in twin studies)
1988–	Linkage analysis	Polygenic inheritance of autism (identification of different locus susceptibility)
2003	DNA sequencing of human genome	Completion of human genome project and sequence of human genomic DNA
2007–	Next generation technology applications	Common variants are the major part of mendelian disorders, autism and other complex disorders
2007–	Epigenetic studies	Understanding the mechanisms of epigenetic regulation in the pathophysiology of autism
2009–	Whole exome-whole genome sequencing	Detection of <i>de novo</i> or inherited mutations in the coding regions and new candidate genes for ASD

ASD: autism spectrum disorders

nant role in the etiology of autism.¹⁹ This review briefly emphasizes autism genetics, examining recent evidences by demonstrating the potential usage of next generation sequencing technologies, particularly WES studies in this respect.

DEVELOPMENT OF NEW SEQUENCING TECHNOLOGIES

In the 1970s, Sanger et al.²⁰ and Maxam and Gilbert²¹ developed new methods to sequence DNA by chain termination and fragmentation techniques, respectively. The technique developed by Sanger and colleagues, commonly referred as Sanger sequencing, required less handling of toxic chemicals and radioisotopes than Maxam and Gilbert's method. As a result, it became the prevailing DNA sequencing method for the next 30 years.^{22,23} Even union of chain termination sequencing by Sanger et al. and the polymerase chain reaction (PCR) by Mullis et al. established many marked events in genetic studies, eventually the completion of Human Genome Project, as well as other genome projects of many species.^{22,24–26}

There is a great diversity of DNA variation in human genome and these variations comprise small base changes (substitutions), rearrangements such as inversions and translocations, insertions and deletions of DNA and large genomic deletions of exons or whole genes. Traditional sequencing methods, especially Sanger sequencing, is restricted to the discovery of variations including small insertions or deletions. For the other remaining mutations dedicated assays are frequently performed with fluorescence in situ hybridization (FISH) for conventional karyotyping, or comparative genomic hybridisation (CGH) microarrays to detect submicroscopic chromosomal CNVs such as microdeletions or duplications.²⁷

Developments between 1977 and 2005 were focused on increasing throughput, accuracy and decreasing cost. Automated Sanger sequencing method defined as the first-generation technology, while Next Generation Sequencing (NGS) repre-

sents the second generation technology. The need for high-throughput, low-cost sequencing drove the development of massively parallel technologies. As a result, new improvements for sequencing technologies developed in recent years.²⁸ NGS libraries in a cell free system were the first major difference. Thousands-to-many-millions of sequencing reactions are produced in parallel which is called “massively parallel sequencing.”²⁹ Also, the sequencing output is directly detected without the need for electrophoresis; base interrogation is performed cyclically and in parallel.^{22,25} The huge number of reads generated by NGS enabled the sequencing of whole genomes with an unprecedented speed and thus it became widely used in various fields of life sciences.²⁵ The first NGS technology to be released in 2005 was the pyrosequencing method by Roche 454 Life Sciences. There are other sequencing platforms including Illumina, Solex, SOLID.^{22,30–32} Third-generation sequencing technology can determine the base composition of single DNA molecules and also enables real-time sequencing.²⁵

NGS involves the whole of genome and whole exome sequencing, aim at developing an understanding the variations in human genome, determines the genetic tendencies to diseases and finding out the pharmacogenetical drug responses. With the evolution of NGS technology, costs of all the genome sequencing activities will gradually decrease. NGS includes different applications including RNA sequencing, ChIP-seq, ChIP-chip, whole genome sequencing, whole genome structural variations, mutation detection and carrier screening, determination of hereditary diseases, preparation of DNA libraries, mitochondrial genome sequencing and individual genomics.³³ This new generation DNA sequencing technology has provided certain benefits in terms of acquiring information pertaining to the genetic/epigenetic regulatory networks, chromatin structure, nuclear structuring and genome variations.^{22,34} Undoubtedly, these kinds of research activities will be at the forefront in the future years.

NGS like many other techniques has some limitations. Exome

Table 2. The latest identified variations with whole exome sequencing studies in ASD group

Gene symbol	Gene name	Mutation type	References
<i>FOXP1</i>	Forkhead Box P1	Frameshift	52
<i>TGM3</i>	Transglutaminase 3	Missense	52
<i>LAMC3</i>	Laminin, Gamma 3	Missense	52
<i>MYO1A</i>	Myosin IA	3'UTR	52
<i>GPR139</i>	G protein-coupled receptor 139	Missense	52
<i>SCN1A</i>	Sodium channel, voltage-gated, type I, alpha subunit	Missense	52
<i>PLCD1</i>	Phospholipase C, delta 1	Synonymous	52
<i>SYNE1</i>	Spectrin repeat containing, nuclear envelope 1	Missense	52
<i>TAF1L</i>	TAF1 RNA polymerase II, TATA box binding protein (TBP)-associated factor	3'UTR	52
<i>DCTN5</i>	Dynactin 5	3'UTR	52
<i>SLC30A5</i>	Solute carrier family 30 (zinc transporter), member 5	Missense	52
<i>IL1R2</i>	Interleukin 1 receptor, type II	Synonymous	52
<i>AFF4</i>	AF4/FMR2 family, member 4	Synonymous	52
<i>GRIN2B</i>	Glutamate receptor, ionotropic, N-methyl D-aspartate 2B	3' splice	52
<i>EPHB6</i>	EPH receptor B6	Synonymous	52
<i>ARHGAP15</i>	Rho GTPase activating protein 15	Synonymous	52
<i>XIRP1</i>	Xin actin-binding repeat containing 1	Missense	52
<i>CHST5</i>	Carbohydrate (N-acetylglucosamine 6-O) sulfotransferase 5	Synonymous	52
<i>TTN</i>	Titin	Synonymous	52
<i>TLK2</i>	Tousled-like kinase 2	Missense	52
<i>RBMS3</i>	RNA binding motif, single stranded interacting protein 3	Missense	52
<i>ADAM33</i>	ADAM metallopeptidase domain 33	Nonsense	53
<i>CSDE1</i>	Cold shock domain containing E1, RNA-binding	Nonsense	53
<i>EPHB2</i>	EPH (Ephrin) receptor B2	Nonsense	53
<i>FAM8A1</i>	Family with sequence similarity 8, member A1	Nonsense	53
<i>FREM3</i>	FRAS1 related extracellular matrix 3	Nonsense	53
<i>MPHOSPH8</i>	M-phase phosphoprotein 8	Nonsense	53
<i>PPM1D</i>	Protein phosphatase 1D magnesium-dependent, delta isoform	Nonsense	53
<i>RAB2A</i>	RAB2A, member RAS oncogene family	Nonsense	53
<i>SCN2A</i>	Sodium channel, voltage-gated, type II, alpha subunit	Nonsense	53
<i>BTN1A1</i>	Butyrophilin, subfamily 1, member A1	Splice Site	53
<i>FCRL6</i>	Fc receptor-like 6	Splice Site	53
<i>KATNAL2</i>	Katanin p60 subunit A-like 2	Splice Site	53
<i>NAPRT1</i>	Nicotinate phosphoribosyltransferase domain containing 1	Splice Site	53
<i>RNF38</i>	Ring finger protein 38	Splice Site	53
<i>SCP2</i>	Sterol carrier protein 2	Frameshift	53
<i>SHANK2</i>	SH3 and multiple ankyrin repeat domains 2	Frameshift	53
<i>ABCA1</i>	ATP-binding cassette, sub-family A (ABC1), member 1	Missense	56
<i>ANK3</i>	Ankyrin 3	Missense	56
<i>CLCN6</i>	Chloride channel, voltage-sensitive 6	Missense	56
<i>HTR3A</i>	5-hydroxytryptamine (serotonin) receptor 3A	Missense	56
<i>RIPK2</i>	Receptor-interacting serine-threonine kinase 2	Missense	56
<i>SLIT3</i>	Slit homolog 3	Missense	56
<i>UNC13B</i>	Uncl3 homolog B	Missense	56

Table 2. The latest identified variations with whole exome sequencing studies in ASD group (continued)

Gene symbol	Gene name	Mutation type	References
<i>FOXP1</i>	Forkhead Box P1	Frameshift	52
<i>TGM3</i>	Transglutaminase 3	Missense	52
<i>LAMC3</i>	Laminin, Gamma 3	Missense	52
<i>MYO1A</i>	Myosin IA	3'UTR	52
<i>GPR139</i>	G protein-coupled receptor 139	Missense	52
<i>SCN1A</i>	Sodium channel, voltage-gated, type I, alpha subunit	Missense	52
<i>PLCD1</i>	Phospholipase C, delta 1	Synonymous	52
<i>SYNE1</i>	Spectrin repeat containing, nuclear envelope 1	Missense	52
<i>TAF1L</i>	TAF1 RNA polymerase II, TATA box binding protein (TBP)-associated factor	3'UTR	52
<i>DCTN5</i>	Dynactin 5	3'UTR	52
<i>SLC30A5</i>	Solute carrier family 30 (zinc transporter), member 5	Missense	52
<i>IL1R2</i>	Interleukin 1 receptor, type II	Synonymous	52
<i>AFF4</i>	AF4/FMR2 family, member 4	Synonymous	52
<i>GRIN2B</i>	Glutamate receptor, ionotropic, N-methyl D-aspartate 2B	3' splice	52
<i>EPHB6</i>	EPH receptor B6	Synonymous	52
<i>ARHGAP15</i>	Rho GTPase activating protein 15	Synonymous	52
<i>XIRP1</i>	Xin actin-binding repeat containing 1	Missense	52
<i>CHST5</i>	Carbohydrate (N-acetylglucosamine 6-O) sulfotransferase 5	Synonymous	52
<i>TTN</i>	Titin	Synonymous	52
<i>TLK2</i>	Tousled-like kinase 2	Missense	52
<i>RBMS3</i>	RNA binding motif, single stranded interacting protein 3	Missense	52
<i>ADAM33</i>	ADAM metalloproteinase domain 33	Nonsense	53
<i>CSDE1</i>	Cold shock domain containing E1, RNA-binding	Nonsense	53
<i>EPHB2</i>	EPH (Ephrin) receptor B2	Nonsense	53
<i>FAM8A1</i>	Family with sequence similarity 8, member A1	Nonsense	53
<i>FREM3</i>	FRAS1 related extracellular matrix 3	Nonsense	53
<i>MPHOSPH8</i>	M-phase phosphoprotein 8	Nonsense	53
<i>PPM1D</i>	Protein phosphatase 1D magnesium-dependent, delta isoform	Nonsense	53
<i>RAB2A</i>	RAB2A, member RAS oncogene family	Nonsense	53
<i>SCN2A</i>	Sodium channel, voltage-gated, type II, alpha subunit	Nonsense	53
<i>BTN1A1</i>	Butyrophilin, subfamily 1, member A1	Splice Site	53
<i>FCRL6</i>	Fc receptor-like 6	Splice Site	53
<i>KATNAL2</i>	Katanin p60 subunit A-like 2	Splice Site	53
<i>NAPRT1</i>	Nicotinate phosphoribosyltransferase domain containing 1	Splice Site	53
<i>RNF38</i>	Ring finger protein 38	Splice Site	53
<i>SCP2</i>	Sterol carrier protein 2	Frameshift	53
<i>SHANK2</i>	SH3 and multiple ankyrin repeat domains 2	Frameshift	53
<i>ABCA1</i>	ATP-binding cassette, sub-family A (ABC1), member 1	Missense	56
<i>ANK3</i>	Ankyrin 3	Missense	56
<i>CLCN6</i>	Chloride channel, voltage-sensitive 6	Missense	56
<i>HTR3A</i>	5-hydroxytryptamine (serotonin) receptor 3A	Missense	56
<i>RIPK2</i>	Receptor-interacting serine-threonine kinase 2	Missense	56
<i>SLIT3</i>	Slit homolog 3	Missense	56
<i>UNC13B</i>	Uncl3 homolog B	Missense	56

Table 2. The latest identified variations with whole exome sequencing studies in ASD group (continued)

Gene symbol	Gene name	Mutation type	References
<i>BRCA2</i>	Breast cancer 2	Missense	55
<i>FAT1</i>	FAT atypical cadherin 1	Missense	55
<i>KCNMA1</i>	Potassium large conductance calcium-activated channel, subfamily M, alpha member 1	Missense	55
<i>CHD8</i>	Chromodomain helicase DNA binding protein 8	Frameshift indel, missense	54
<i>NTNG1</i>	Netrin G1	Missense	54
<i>GRIN2B</i>	Glutamate receptor, ionotropic, N-methyl D-aspartate 2B	Frameshift indel, nonsense, splice site	54
<i>LAMC3</i>	Laminin, Gamma 3	Missense	54
<i>SCN1A</i>	Sodium channel, voltage-gated, type I, alpha subunit	Missense	54
<i>CTTNBP2</i>	Cortactin binding protein 2	Frameshift deletion	58
<i>RIMS1</i>	Regulating synaptic membrane exocytosis 1	Frameshift insertion	58
<i>DYRK1A</i>	Dual-specificity tyrosine-(Y)-phosphorylation regulated kinase 1A	Frameshift deletion	58
<i>ZFYVE26</i>	Zinc finger, FYVE domain containing 26	Frameshift deletion	58
<i>DST</i>	Dystonin	Frameshift deletion	58
<i>ANK2</i>	Ankyrin 2	Nonsense	58
<i>UBE3B</i>	Ubiquitin protein ligase E3B	Missense	57
<i>CLTCL1</i>	Clathrin, heavy chain-like 1	Missense	57
<i>ZNF18</i>	Zinc finger protein 18	Missense	57
<i>AMT</i>	Aminomethyltransferase	Missense	62
<i>PEX7</i>	Peroxisomal biogenesis factor 7	Missense	62
<i>SYNE1</i>	Spectrin repeat containing, nuclear envelope 1	Missense	62
<i>VPS13B</i>	Vacuolar protein sorting 13 homolog B	Missense, frameshift	62
<i>PAH</i>	Phenylalanine hydroxylase	Frameshift	62
<i>POMGNT1</i>	Protein O-linked mannose N-acetylglucosaminyltransferase 1	Missense	62
<i>ANK3</i>	Ankyrin 3	Missense	61
<i>CIC</i>	Capicua Transcriptional Repressor	Missense	47
<i>GLUD2</i>	Glutamate dehydrogenase 2	Missense	47
<i>ROGD1</i>	Rogdi homolog (Drosophila)	Missense	47
<i>SEZ6</i>	Seizure related 6 homolog	Missense	47
<i>CEP290</i>	Centrosomal protein 290 kDa	Missense	47
<i>CSMD1</i>	CUB and Sushi multiple domains 1	Missense	47
<i>FAT1</i>	FAT atypical cadherin 1	Missense	47
<i>STXBP5</i>	Syntaxin binding protein 5	Missense	47
<i>CHD2</i>	Chromodomain helicase DNA binding protein 2	Frameshift	64
<i>KMT2E</i>	Lysine (K)-specific methyltransferase 2E	Frameshift	64
<i>PHF3</i>	PHD finger protein 3	Frameshift	64
<i>RIMS1</i>	Regulating synaptic membrane exocytosis 1	Frameshift	64
<i>KCND2</i>	Potassium voltage-gated channel, Shal-related subfamily, member 2	Missense	66
<i>BICC1</i>	BicC family RNA binding protein 1	Missense	66
<i>SLC8A2</i>	Solute carrier family 8 (sodium/calcium exchanger), member 2	Missense	66
<i>GPR124</i>	G protein-coupled receptor 124	Missense	66

Table 2. The latest identified variations with whole exome sequencing studies in ASD group (continued)

Gene symbol	Gene name	Mutation type	References
<i>IL1RAPL1</i>	Interleukin 1 Receptor Accessory Protein-Like 1	Missense	68
<i>GPRASP2</i>	G protein-coupled receptor associated sorting protein 2	Missense	68
<i>GABRQ</i>	Gamma-aminobutyric acid (GABA) A receptor, theta	Missense	68
<i>SYTL4</i>	Synaptotagmin-like 4	Missense	68
<i>PIR</i>	Pirin (iron-binding nuclear protein)	Missense	68
<i>CHAC1</i>	Cation Transport Regulator Homolog 1	Frameshift	68
<i>RPS24</i>	Ribosomal protein S24	Nonsense	67
<i>CD300LF</i>	CD300 molecule-like family member f	Frameshift	67
<i>CDKL5</i>	Cyclin-Dependent Kinase-Like 5	Missense	69
<i>SCN2A</i>	Sodium channel, voltage-gated, type II, alpha subunit	Nonsense	69
<i>CUL3</i>	Cullin 3	Missense	69
<i>MED13L</i>	Mediator complex subunit 13-like	Frameshift	69
<i>KCNV1</i>	Potassium channel, subfamily V, member 1	Frameshift	69
<i>MAOA</i>	Monoamine oxidase A	Splice site	69
<i>PTEN</i>	Phosphatase and tensin homolog	Splice site	69
<i>FOXP1</i>	Forkhead box protein P1	Frameshift	70
<i>SLC7A11</i>	Solute carrier family 7 member 11	Missense	71
<i>ICA1</i>	Islet cell autoantigen 1	Missense	71
<i>DNAJC1</i>	DnaJ (Hsp40) homolog, subfamily C, member 1	Missense	71
<i>C1S</i>	Complement component 1, s subcomponent	Missense	71
<i>TRAPPC12</i>	Trafficking protein particle complex 12	Missense	71
<i>CLN8</i>	Ceroid-Lipofuscinosis, Neuronal 8	Missense	71

ASD: autism spectrum disorders

enrichment is the basis for exome sequencing technology. The exome regions are enriched by different methods including hybridization capture or solution based methodology.³⁵ Differently from Sanger sequencing, each run of samples generated a single sequencing read. The specific location of each read must be determined computationally referred as mapping or alignment. Secondly, multiple coverage is required to analyze the full allelic content of the sample.²⁹ Fragment sequencing has higher sequencing error rates than Sanger sequencing; thus, further validation of identified variations using Sanger sequencing is very important.³⁶ The weaknesses of these technologies are solved with the arrival of third-generation sequencing technologies.³⁵ HapMap or dbSNP databases, cannot effectively exclude common and irrelative variants that obtained from WES. With the application of the Bonferroni correction and true bioinformatic approaches, the true variants would be distinguished through the yielded threshold.³⁷ Statistical metrics are essential to distinguish false positive data from true positives.³⁸ The researchers should be focused on the non-synonymous (NS) variants, splice acceptor-site or donor-site mutations, and frameshift mutations (insertions/deletions).

Whole exome sequencing in Mendelian disorders and complex diseases

Determination of the genetic basis which lies behind the rare single gene diseases is important for understanding to the mechanism of disease in terms of finding out its role in the biological pathways and developing the treatment.³⁷ Since researches on wide genome linkage and Mendelian inheritance pattern (autosomal dominant, autosomal recessive, X-linked recessive) while they demonstrate a perfect segregation, because of this, successful results were obtained in the determination of these variants.^{39,40} Exome sequencing serves for the sequencing of all the exons in the genome. Since a significant part of mutations exists in the exons, this approach is efficient in revealing the Mendelian diseases. Success of exome sequencing in revealing these mutations and identifying genes have been demonstrated by several studies.⁴⁰⁻⁴² It is estimated that 85% of the disease-causing mutations are located in coding and functional regions of the genome.⁴³ Therefore, sequencing of the exome, has the potential to uncover the causes of large number of rare, mostly monogenic (Mendelian disorders), genetic disorders as well as predisposing variations in common diseases.²⁴

Within the last five years, researches have made numerous

studies regarding the complex diseases. Before these researches, complex phenotype studies mainly focused on the candidate gene, linkage and association studies. Although the linkage analysis were used to define the variants in thousands of Mendelian diseases, this method was not effective in complex diseases due to the effect of many genetic and environmental factors. Additionally, exome sequencing provides new opportunities in complex and rare diseases, and also in sporadic cases, which have been used since 2007.⁴⁴ However, currently only the allelic variants responsible from only less than half of these diseases have been determined. Factors aggravating the diagnosis include the existence of only one case within the family and locus heterogeneity (mutations in different locus result in the same phenotype). Each of these factors limit the strategies used in candidate gene determination.^{39,40} Due to the difficulty in finding sufficient number of samples, linkage studies are not effective in focusing on the rare diseases and sporadic cases. In addition, linkage studies may also fail due to the genetic and phenotypic heterogeneity in Mendelian diseases. Exome sequencing has been used in defining the variants of several rare diseases such as the Kabuki Syndrome, Miller Syndrome and the Schinzel-Giedion Syndrome. WES may be considered as an efficient technique in diseases which demonstrate genetic heterogeneity.^{38,42} GWAS of neuropsychiatric disorders have unequivocally shown that common variants with large effects do not underlie schizophrenia or autism.⁴⁵ For this reason, it is important to use such technologies to discover new molecular pathways which are responsible for these complex disorders.

Whole exome sequencing implications in autism

With the advances in genomic technology, such as array-based GWAS and array comparative genomic hybridization (aCGH), researchers have been able to detect several rare genomic rearrangements and CNVs in subsets of cases with ASD.¹⁸ GWAS and WES have allowed the potential genetic mechanisms underlying the overlap between psychiatric disorders to be investigated more directly and easily.^{18,46} WES seems to be frequently used in complex disorders including autism and the number of such studies is increasing gradually (Table 2).⁴⁷⁻⁵¹ One of the advantages of WES is that a specific gene which harbors a genetic variant can be identified. Additionally, related biological pathways can be further investigated. The function of specific types of mutation, including nonsense and missense mutations can be predicted by the help of publicly available bioinformatic resources.¹⁴

The exomes of 20 individuals with sporadic ASD patients and their parents (trio-based study) were sequenced. Totally, 21 *de novo* mutations were identified and 11 of these mutations were altering the protein structure. The researchers sug-

gested that trio-based exome sequencing is a powerful approach for identifying new candidate genes for ASD.⁵² Sanders et al.⁵³ showed a total of 279 identified *de novo* coding mutations using WES of 928 individuals. Interestingly, two independent nonsense variants disrupt the same *SCN2A* gene. A total of 677 individual exomes from 209 families were sequenced in 2012. Moreover, 39% (49 of 126) of the most severe *de novo* mutations were related with a highly interconnected β -catenin/chromatin remodelling protein network as new candidate genes for autism. In probands' exomes, protein-altering mutations were observed in *CHD8* and *NTNG1* genes.⁵⁴ In another study, Neale et al.⁵⁵ assessed the role of *de novo* mutations by sequencing the exomes of ASD cases and their parents (n=175 trios). They observed totally 161 coding region point mutations (50 silent, 101 missense, and 10 nonsense), 2 conserved splice site (CSS) single nucleotide variations (SNVs) and 6 frameshift indels. Their results provided strong evidence in favor of *CHD8* and *KATNAL2* genes were likely to be important genetic risk factors. In the same year another WES study was conducted in a cohort of 20 ASD patients. They also sequenced an additional 47 ASD samples, and identified three different missense mutations in *ANK3* gene in four unrelated ASD cases. One of the mutation (c.4705T>G/p.S1569A) is a *de novo* mutation. With this finding the authors suggested an association between *ANK3* mutations and ASD susceptibility and imply a shared molecular pathophysiology between ASD and other neuropsychiatric disorders such as schizophrenia, and bipolar disorders. *ANK3* gene is a member of the ankyrin family of proteins that is associated with the spectrin-actin cytoskeleton in neuronal cells. Loss of function of *ANK3* may influence neuronal excitability through ion channel function and affects synaptic development and functions.⁵⁶

WES of 16 probands by the completion of homozygosity mapping revealed validated homozygous, potentially pathogenic recessive mutations that segregated perfectly with disease in 4 families. Identified mutations except in the *NCKAP5L* gene (the other candidate genes were *UBE3B*, *CLTCL1* and *ZNF18*) were probably or possibly damaging according to PolyPhen 2. The candidate genes that identified in this study (*UBE3B*, *CLTCL1*, *NCKAP5L*, *ZNF18*) encode proteins involved in different pathways especially proteolysis, GTPase-mediated signaling, cytoskeletal organization.⁵⁷ Exome sequencing of simplex 343 families with ASD revealed *de novo* small indels and point mutations. This study demonstrated that gene-disrupting mutations (nonsense, splice site, and frame shift) are twice as frequent in affected versus unaffected children.⁵⁸ In another study a large autism family with five generation (47 family members) were investigated with both WES and linkage analysis. The authors obtained strong asso-

ciation for localization of a risk locus to chromosome 22, precisely bound the interval likely to carry the risk variant, and prioritize evaluation of all exome sequence variants within that region.⁵⁹ Exome sequencing was applied to the X chromosome in 12 unrelated families with two affected males with a different approach. A nonsense mutation in the *TMLHE* gene was identified in two brothers with autism and intellectual disability. Further functional analyses confirmed that the mutations were associated with a loss-of-function and this finding supported the rare variants on the X chromosome are involved in the etiology of ASD.⁶⁰ With WES, two probands from a large pedigree including two parents and eight children examined. The researchers identified 59 candidate variants that may increase susceptibility to autism. They suggested only one gene, *ANK3* (c.11068G>A/p.G3690R mutation), as the most likely candidate gene in this family.⁶¹ WES applied to 11 ASD families enriched for inherited causes due to consanguinity and this study determined familial ASD associated with biallelic mutations in disease genes (*PAH*, *SYNE1*, *AMT*, *POMGNT1*, *PEX7*, *VPS13B*), some of them implicated for the first time in ASD. *PEX7* gene encodes a receptor required for import of PTS2 (peroxisome targeting signal 2)-containing proteins into the peroxisome.⁶²

Total of 42 Australian ASD families with 48 probands were screened with WES. Among 44 *de novo* variants, there were 4 were intergenic, 29 associated with protein-coding sequences, 9 were intronic, and 2 occur in 3'-UTR regions of genes. Gene ontology analysis revealed that identified *de novo* variants cluster in key neurobiological processes involving neuronal development, signal transduction and synapse development including the neurexin trans-synaptic complex.⁶³ Cukier et al.⁴⁸ performed WES on 164 individuals from 40 families with multiple affected individuals. According to their data, potentially novel ASD candidates identified that including genes with previous clinical and molecular evidence supporting a neuronal function such as *RODG1*, *CIC*, *NTSR2*, *GLUD2* and *SEZ6*. Additional variations were also found in previously reported autism candidate genes including *CEP290*, *CSMD1*, *FAT1*, and *STXBP5*. WES data from 787 ASD families (2,963 individuals), two ASD-associated genes, *KMT2E* (*MLL5*), a chromatin regulator, and *RIMS1*, a regulator of synaptic vesicle release were identified in this study.⁶⁴ WES was performed with 488 ASD cases and 372 controls of European ancestry in 2014. The authors identified variant in one gene, Fanconi-associated nuclease 1 (*FAN1*) has being associated with schizophrenia and ASD. They suggested that *FAN1* is a key driver in the 15q13.3 locus, encoding a DNA repair enzyme, for the neurodevelopmental and associated psychiatric phenotypes.⁶⁵ In the same year another study of WES was published on a family with identical twins affected with autism and intractable

seizures. A *de novo* variant was identified in the *KCND2* gene which encodes the Kv4.2 potassium channel and mutant protein disrupts potassium current inactivation. This alteration (p.Val404Met) is novel and occurs in a highly conserved region.⁶⁶ In a recent study where two rare heterozygous truncating variations, Q191X in *RPS24* and P261fsX266 in *CD300LF* have been identified as risk candidates for ASD. *RPS24* gene encodes ribosomal protein 24 (RPS24), which is involved in ribosome biogenesis. CD300LF is a member of the CD300 gene complex and encodes CMRF35-like molecule 1 (CLM-1), which is an immunoreceptor expressed on myeloid cells. CLM-1 overexpression reduced acute brain injury therefore CLM-1 has protective effects against neuroinflammation.⁶⁷ Thirty females with autism from multiplex families have been selected to WES case study. Functional variants of X-linked genes (*GABRQ*, *IL1RAPL1*, *PIR*, *GPRASP2* and *SYTL4*) were identified in 5 females.⁶⁸ The latest study with exome sequencing was conducted with 36 males with a diagnosis of idiopathic ASD. Five *de novo* (*SCN2A*, *MED13L*, *KCNV1*, *CUL3*, and *PTEN*) and two inherited X-linked variants (*MAOA* and *CDKL5*) were identified in seven cases. *CDKL5* gene is a member of the Ser/Thr protein kinase family and encodes a phosphorylated protein with protein kinase activity. *PTEN* gene is a part of the PI3K-Akt and PDGF signaling pathways. *SCN2A* gene is involved in axon guidance and L1CAM interaction pathways.⁶⁹ A heterozygous *de novo* *FOXP1* variant (c.1267_1268delGT, p.V423Hfs*37) was identified with exome sequencing in a patient with autism, intellectual disability and severe speech and language impairment. This variant disrupts *FOXP1* activity, including subcellular localization and transcriptional repression properties.⁷⁰ Egawa et al.⁷¹ performed WES in two families with three affected siblings. Six novel missense variants were determined with this study including *SLC7A11*, *ICA1*, *DNAJC1*, *CIS*, *TRAPPC12* and *CLN8* genes. They suggested that *CLN8* is a potential genetic risk factor for ASD and this gene plays a role in cell proliferation during neuronal differentiation and in protection against neuronal cell apoptosis.

CONCLUSION

The arrival of NGS and WES methods comprise the beginning of a new era, not just for autism research but also for researches into nearly every complex disorder. With new sequencing data, scientists will benefit greatly from the ability to combine genetic data generated by multiple methods (for example, combining GWAS and/or linkage data with WES data). WGS will become increasingly more cost-effective as sequencing costs decrease and bioinformatics tools improve, and these

advances will open up the possibility of detecting non-coding genetic changes in regulatory regions in the genome.¹⁴ The WES approach has some limitations as it only detects individual genetic variations based on single nucleotide polymorphisms (SNPs) and small indels associated with exonic regions of the genome. WES technology does not detect large CNV or genetic variants of regulatory sequences located in intergenic regions. Compared to traditional methods, exome sequencing is a less costly and a faster technique for the determination of mutations as well as for autism and in other neurological diseases where lots of candidate genes and loci are responsible from these disorders. WES allows scanning the mutations/ variations in genes commonly observed in many diseases with genetic heterogeneity.⁷² As a result, it is an undeniable fact that WES will also be used as a golden standard for more complex diseases as well as ASD and be used with all genome sequencing activities, along with their use for revealing the underlying genes of the rare genetic disorders in the future.

REFERENCES

- Devlin B, Scherer SW. Genetic architecture in autism spectrum disorder. *Curr Opin Genet Dev* 2012;22:229-237.
- Siniscalco D, Cirillo A, Bradstreet JJ, Antonucci N. Epigenetic findings in autism: new perspectives for therapy. *Int J Environ Res Public Health* 2013;10:4261-4273.
- Chung BH, Tao VQ, Tso WW. Copy number variation and autism: new insights and clinical implications. *J Formos Med Assoc* 2014;113:400-408.
- Sener EF, Oztop DB, Ozkul Y. MTHFR gene C677T polymorphism in autism spectrum disorders. *Genet Res Int* 2014;2014:698574.
- Srivastava AK, Schwartz CE. Intellectual disability and autism spectrum disorders: causal genes and molecular mechanisms. *Neurosci Biobehav Rev* 2014;46:161-174.
- Canitano R. Epilepsy in autism spectrum disorders. *Eur Child Adolesc Psychiatry* 2007;16:61-66.
- Liu X, Takumi T. Genomic and genetic aspects of autism spectrum disorder. *Biochem Biophys Res Commun* 2014;452:244-253.
- Smith CL, Bolton A, Nguyen G. Genomic and epigenomic instability, fragile sites, schizophrenia and autism. *Curr Genomics* 2010;11:447-469.
- Grabrucker AM. Environmental factors in autism. *Front Psychiatry* 2013;3:118.
- Goldani AA, Downs SR, Widjaja F, Lawton B, Hendren RL. Biomarkers in autism. *Front Psychiatry* 2014;5:100.
- Tordjman S, Somogyi E, Coulon N, Kermarrec S, Cohen D, Bronsard G, et al. Gene×Environment interactions in autism spectrum disorders: role of epigenetic mechanisms. *Front Psychiatry* 2014;5:53.
- Sener EF. Association of copy number variations in autism spectrum disorders: a systematic review. *Chinese J Biol* 2014;2014:1-9.
- Schanen NC. Epigenetics of autism spectrum disorders. *Hum Mol Genet* 2006;15:R138-R150.
- Schreiber M, Dorschner M, Tsuang D. Next-generation sequencing in schizophrenia and other neuropsychiatric disorders. *Am J Med Genet B Neuropsychiatr* 2013;162B:671-678.
- Kato T. Whole genome/exome sequencing in mood and psychotic disorders. *Psychiatry Clin Neurosci* 2015;69:65-76.
- Han G, Sun J, Wang J, Bai Z, Song F, Lei H. Genomics in neurological disorders. *Genomics Proteomics Bioinformatics* 2014;12:156-163.
- Rosti RO, Sadek AA, Vaux KK, Gleeson JG. The genetic landscape of autism spectrum disorders. *Dev Med Child Neurol* 2014;56:12-18.
- Doherty JL, Owen MJ. Genomic insights into the overlap between psychiatric disorders: implications for research and clinical practice. *Genome Med* 2014;6:29.
- Toma C, Torrico B, Hervás A, Valdés-Mas R, Tristán-Noguero A, Padillo V, et al. Exome sequencing in multiplex autism families suggests a major role for heterozygous truncating mutations. *Mol Psychiatry* 2014;19:784-790.
- Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA* 1977;74:5463-5467.
- Maxam AM, Gilbert W. A new method for sequencing DNA. *Proc Natl Acad Sci U S A* 1977;74:560-564.
- van Dijk EL, Auger H, Jaszczyszyn Y, Thermes C. Ten years of next-generation sequencing technology. *Trends Genet* 2014;30:418-426.
- França LT, Carrilho E, Kist TB. A review of DNA sequencing techniques. *Q Rev Biophys* 2002;35:169-200.
- Rabbani B, Tekin M, Mahdieh N. The promise of whole-exome sequencing in medical genetics. *J Hum Genet* 2014;59:5-15.
- Yang Y, Xie B, Yan J. Application of next-generation sequencing technology in forensic science. *Genomics Proteomics Bioinformatics* 2014;12:190-197.
- International Human Genome Consortium. Finishing the euchromatic sequence of the human genome. *Nature* 2004;431:931-945.
- Behjati S, Tarpey PS. What is next generation sequencing? *Arch Dis Child Educ Pract Ed* 2013;98:236-238.
- Day-Williams AG, Zeggini E. The effect of next-generation sequencing technology on complex trait research. *Eur J Clin Invest* 2011;41:561-567.
- Moorthie S, Mattocks CJ, Wright CF. Review of massively parallel DNA sequencing technologies. *Hugo J* 2011;5:1-12.
- Xuan J, Yu Y, Qing T, Guo L, Shi L. Next-generation sequencing in the clinic: promises and challenges. *Cancer Lett* 2013;340:284-295.
- Kayman Kürekçi G, Dinçer P. Next-generation DNA sequencing technologies. *Erciyes Med J* 2014;36:99-103.
- Shendure J, Ji H. Next-generation DNA sequencing. *Nat Biotechnol* 2008;26:1135-1145.
- Pareek CS, Smoczynski R, Tretyn A. Sequencing technologies and genome sequencing. *J Appl Genet* 2011;52:413-435.
- Desai AN, Jere A. Next-generation sequencing: ready for the clinics? *Clin Genet* 2012;81:503-510.
- Zhang X. Exome sequencing greatly expedites the progressive research of Mendelian diseases. *Front Med* 2014;8:42-57.
- Ku CS, Wu M, Cooper DN, Naidoo N, Pawitan Y, Pang B, et al. Exome versus transcriptome sequencing in identifying coding region variants. *Expert Rev Mol Diagn* 2012;12:241-251.
- Ng SB, Buckingham KJ, Lee C, Bigham AW, Tabor HK, Dent KM, et al. Exome sequencing identifies the cause of a mendelian disorder. *Nat Genet* 2010;42:30-35.
- Biesecker LG. Exome sequencing makes medical genomics a reality. *Nat Genet* 2010;42:13-14.
- Wang JL, Yang X, Xia K, Hu ZM, Weng L, Jin X, et al. TGM6 identified as a novel causative gene of spinocerebellar ataxias using exome sequencing. *Brain* 2010;133:3510-3518.
- Hedges D, Burges D, Powell E, Almonte C, Huang J, Young S, et al. Exome sequencing of a multigenerational human pedigree. *PLoS One* 2009;4:e8232.
- Kuhlenbäumer G, Hullmann J, Appenzeller S. Novel genomic techniques open new avenues in the analysis of monogenic disorders. *Hum Mutat* 2011;32:144-151.
- Ng SB, Bigham AW, Buckingham KJ, Hannibal MC, McMillin MJ, Gildersleeve HI, et al. Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome. *Nat Genet* 2010;42:790-793.
- Majewski J, Schwartzentruber J, Lalonde E, Montpetit A, Jabado N. What can exome sequencing do for you? *J Med Genet* 2011;48:580-589.
- Ku CS, Naidoo N, Pawitan Y. Revisiting Mendelian disorders through exome sequencing. *Hum Genet* 2011;129:351-370.
- Ezewudo M, Zwick ME. Evaluating rare variants in complex disorders

- using next-generation sequencing. *Curr Psychiatry Rep* 2013;15:349.
46. Jouan L, Gauthier J, Dion PA, Rouleau GA. Rare variants in complex traits: novel identification strategies and the role of de novo mutations. *Hum Hered* 2012;74:215-225.
 47. De Rubeis S, He X, Goldberg AP, Poultney CS, Samocha K, Cicek AE, et al. Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature* 2014;515:209-215.
 48. Cukier HN, Dueker ND, Slifer SH, Lee JM, Whitehead PL, Lalanne E, et al. Exome sequencing of extended families with autism reveals genes shared across neurodevelopmental and neuropsychiatric disorders. *Mol Autism* 2014;5:1.
 49. Lee H, Lin MC, Kornblum HI, Papazian DM, Nelson SF. Exome sequencing identifies de novo gain of function missense mutation in KCND2 in identical twins with autism and seizures that slows potassium channel inactivation. *Hum Mol Genet* 2014;23:3481-3489.
 50. Fukai R, Hiraki Y, Yofune H, Tsurusaki Y, Nakashima M, Saito H, et al. A case of autism spectrum disorder arising from a de novo missense mutation in POGZ. *J Hum Genet* 2015;60:277-279.
 51. Yuen RK, Thiruvahindrapuram B, Merico D, Walker S, Tammimies K, Hoang N, et al. Whole-genome sequencing of quartet families with autism spectrum disorder. *Nat Med* 2015;21:185-191.
 52. O'Roak BJ, Deriziotis P, Lee C, Vives L, Schwartz JJ, Girirajan S, et al. Exome sequencing in sporadic autism spectrum disorders identifies severe de novo mutations. *Nature Genet* 2011;43:585-589.
 53. Sanders SJ, Murtha MT, Gupta AR, Murdoch JD, Raubeson MJ, Willsey AJ, et al. De novo mutations revealed by whole exome sequencing are strongly associated with autism. *Nature* 2012;485:237-241.
 54. O'Roak BJ, Vives L, Girirajan S, Karakoc E, Krumm N, Coe BP, et al. Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. *Nature* 2012;485:246-250.
 55. Neale BM, Kou Y, Liu L, Ma'ayan A, Samocha KE, Sabo A, et al. Patterns and rates of exonic de novo mutations in autism spectrum disorders. *Nature* 2012;485:242-245.
 56. Bi C, Wu J, Jiang T, Liu Q, Cai W, Yu P, et al. Mutations of ANK3 identified by exome sequencing are associated with autism susceptibility. *Hum Mutat* 2012;33:1635-1638.
 57. Chahrour MH, Yu TW, Lim ET, Ataman B, Coulter ME, Hill RS, et al. Whole-exome sequencing and homozygosity analysis implicate depolarization-regulated neuronal genes in autism. *PLoS Genet* 2012;8:e1002635.
 58. Iossifov I, Ronemus M, Levy D, Wang Z, Hakker I, Rosenbaum J, et al. De novo gene disruptions in children on the autistic spectrum. *Neuron* 2012;74:285-299.
 59. Marchani EE, Chapman NH, Cheung CY, Ankenman K, Stanaway IB, Coon HH, et al. Identification of rare variants from exome sequence in a large pedigree with autism. *Hum Hered* 2012;74:153-164.
 60. Nava C, Lamari F, Héron D, Mignot C, Rastetter A, Keren B, et al. Analysis of the chromosome X exome in patients with autism spectrum disorders identified novel candidate genes, including TMLHE. *Transl Psychiatry* 2012;2:e179.
 61. Shi L, Zhang X, Golhar R, Otieno FG, He M, Hou C, et al. Whole-genome sequencing in an autism multiplex family. *Mol Autism* 2013;4:8.
 62. Yu TW, Chahrour MH, Coulter ME, Jiralerspong S, Okamura-Ikeda K, Ataman B, et al. Using whole-exome sequencing to identify inherited causes of autism. *Neuron* 2013;77:259-273.
 63. An JY, Cristino AS, Zhao Q, Edson J, Williams SM, Ravine D, et al. Towards a molecular characterization of autism spectrum disorders: an exome sequencing and systems approach. *Transl Psychiatry* 2014;4:e394.
 64. Dong S, Walker MF, Carriero NJ, DiCola M, Willsey AJ, Ye AY, et al. De novo insertions and deletions of predominantly paternal origin are associated with autism spectrum disorder. *Cell Rep* 2014;9:16-23.
 65. Ionita-Laza I, Xu B, Makarov V, Buxbaum JD, Roos JL, Gogos JA, et al. Scan statistic-based analysis of exome sequencing data identifies FAN1 at 15q13.3 as a susceptibility gene for schizophrenia and autism. *Proc Natl Acad Sci U S A* 2014;111:343-348.
 66. Lee H, Lin MC, Kornblum HI, Papazian DM, Nelson SF. Exome sequencing identifies de novo gain of function missense mutation in KCND2 in identical twins with autism and seizures that slows potassium channel inactivation. *Hum Mol Genet* 2014;23:3481-3489.
 67. Inoue E, Watanabe Y, Egawa J, Sugimoto A, Nunokawa A, Shibuya M, et al. Rare heterozygous truncating variations and risk of autism spectrum disorder: whole-exome sequencing of a multiplex family and follow-up study in a Japanese population. *Psychiatry Clin Neurosci* 2015;69:472-476.
 68. Butler MG, Rafi SK, Hossain W, Stephan DA, Manzardo AM. Whole exome sequencing in females with autism implicates novel and candidate genes. *Int J Mol Sci* 2015;16:1312-1335.
 69. Codina-Solà M, Rodríguez-Santiago B, Homs A, Santoyo J, Rigau M, Aznar-Lain G, et al. Integrated analysis of whole-exome sequencing and transcriptome profiling in males with autism spectrum disorders. *Mol Autism* 2015;6:21.
 70. Lozano R, Vino A, Lozano C, Fisher SE, Deriziotis P. A de novo FOXP1 variant in a patient with autism, intellectual disability and severe speech and language impairment. *Eur J Hum Genet* 2015;23:1702-1707.
 71. Egawa J, Watanabe Y, Wang C, Inoue E, Sugimoto A, Sugiyama T, et al. Novel rare missense variations and risk of autism spectrum disorder: whole-exome sequencing in two families with affected siblings and a two-stage follow-up study in a Japanese population. *PLoS One* 2015;10:e0119413.
 72. Singleton AB. Exome sequencing: a transformative technology. *Lancet Neurol* 2011;10:942-946.