AUTISM SPECTRUM DISORDER - A COMPLEX GENETIC DISORDER

Hristo Y. Ivanov1, Vili K. Stoyanova1,2, Nikolay T. Popov3, Tihomir I. Vachev1,4
1Department of Pediatrics and Medical Genetics, Medical Faculty, Medical University, Plovdiv; 2Department of Medical Genetics, St. George University Hospital, Plovdiv; 3Psychiatric Ward for Men, State Psychiatric Hospital, Pazardzhik; 4Department of Plant Physiology and Molecular Biology, Paisii Hilendarski University, Plovdiv, Bulgaria

ABSTRACT

Autism spectrum disorder is an entity that reflects a scientific consensus that several previously separated disorders are actually a single spectrum disorder with different levels of symptom severity in two core domains - deficits in social communication and interaction, and restricted repetitive behaviors. Autism spectrum disorder is diagnosed in all racial, ethnic and socioeconomic groups and because of its increased prevalence, reported worldwide through the last years, made it one of the most discussed child psychiatric disorders. In term of aetiology as several other complex diseases, Autism spectrum disorder is considered to have a strong genetic component.

Key words: autism spectrum disorder, whole exome sequencing, genome wide expression

INTRODUCTION

The term and the diagnosis of ‘infantile autism’ were introduced into the world of science by an American psychiatrist, Leo Kanner.1 In an article of 1943, Kanner described a condition in 11 children that he called ‘early infantile autism’.
failed to demonstrate empathy, did not understand body language and had poor language in terms of sounds or verbal statements. He gave a description of a syndrome which was later named after him, the syndrome of Asperger.

Over the past 65 years etiological paradigms within psychiatry have changed significantly in parallel with the development of the concept that cognitive and behavioral disorders have organic ‘brain-based’ etiology. Initially, autism was thought to be largely a condition resulting from the way children were brought up by their parents. The first suggestion that rare genetic and biomedical conditions can lead to autism and that it is not associated with ‘bad parenting’, was voiced by phenylketonuria, but then these syndromic cases of autism were considered exceptions.\(^2,3\) In the 60s and 70s of the last century autism was considered as a form of psychosis or childhood schizophrenia. The separation of autism from childhood-onset psychoses, particularly schizophrenia, was an important advance in the study of childhood psychopathology. Clinically, this has been a basic and self evident distinction as age of onset, differential diagnosis, and treatment of the two conditions differed.\(^4\) In 1980s autism was classified as a developmental disorder and the thesis of its biological nature was accepted. The term ‘pervasive developmental disorders’ (PDD) was used and introduced into the Diagnostic and Statistic Manual of Mental Disorders - III (DSM - III) and later included in the International Classification of Diseases - 10 (ICD). In the next version of DSM, fourth edition (DSM-IV)\(^5\), PDD was conceptualized as a manifestation of triad symptoms: impairment in social interaction, qualitative impairment of communication, restricted and repetitive behavior pattern, and included five specific subgroups: Autistic Disorder; Pervasive Developmental Disorder, Not Otherwise Specified; Asperger’s Disorder; Rett’s Disorder; and Childhood Disintegrative Disorder. In the newest version of DSM-V released in 2013\(^6\), these subgroups were merged into a single umbrella term as ‘autism spectrum disorders’ (ASD), and the triad of impairments has been folded into two, with social communication and social interaction combined as a single diagnostic domain.

**PREVALENCE**

ASD is diagnosed in all racial, ethnic and socio-economic groups, five times more common in boys than in girls.\(^7\)

Prior to 1990, most studies estimated general population prevalence for autism of four to five per 10,000 (1/2000-1/2500).\(^8\) Increased prevalence of ASD have been reported worldwide over the last few years. According to the Center for Disease Control (CDC) and Autism and Developmental Disabilities Monitoring (ADDM) Network in the USA, 1 in 88 children are identified with the ASD\(^7\) although this increase might merely reflect improved diagnostic means.

Due to the high social significance of autism and the arising frequency, clinical practice has long sought to shade a light on etiology and pathogenesis of this disorder which still remain elusive. Until the etiology of ASD is known, it will be difficult to identify effective prevention and treatment. Current belief holds that ASD derive generically from a complex interaction between genetic, epigenetic and environmental factors\(^9\) (Fig. 1).

Further, it is suggested that some of the associated sequence variations detected in ASD are common in the general population. However, it is not clear whether the involvement of single genes in combination with nongenetic factors or multiple genes through locus heterogeneity or multiple genes through allelic heterogeneity is needed to result in the ASD phenotype. It is also suggested that multiple genes in combination with nongenetic factors may be in operation and that is needed to result in the ASD phenotype.\(^10\) An important fact to note is that environmental factors are probably connected somehow to genetic makeup and epigenetic mechanisms and the non-coding RNAs can be the link between, even though there is no clear evidence for that. All those questions still remain unclear.\(^11\)

**THE ENVIRONMENTAL MODIFIERS**

Environmental exposures and stressors act through their impact on the organism. Effects may impact brain development, development of other organs and systems, and ongoing physiological processes.\(^12\) Such environmental modifiers include advanced parental age, oxidative stress, neuroinflammation and mitochondrial dysfunction as well as environmental pollutants (e.g., air pollution, organophosphates and heavy metals) and biochemical disturbances.\(^10\) Other modulating factors include an interaction between immune dysfunction and genetic predisposition, as evident from the reports of cytokine-mediated influences on neuronal development and links between genes that encode for immune-related proteins and ASD.\(^13\)

As there is controversy, a recent study showed
Autism Spectrum Disorder - a Complex Genetic Disorder

that part of the individual risk of ASD and autistic disorder increased with increasing genetic relatedness. Heritability of ASD and autistic disorder was estimated in this study to be approximately 50%.

Environmental factors do contribute to the etiology of ASD but they are beyond the scope of the present review, and we shall concentrate on the genetic basis of ASD.

ASD has a strong genetic basis. In twin studies, the concordance rate of broad ASD phenotype in monozygotic twins is 70–90%, but 0–30% in dizygotic twins. A very recent prospective study reported that if there was an older sibling with ASD, the likelihood of subsequent offspring having autism was 18.7% overall and 32.2% if the child had more than one older sibling with autism. Despite being highly heritable, ASD is genetically complex and the underlying genetic architecture is still not well understood.

THE TERMS ‘SYNDROMIC’ AND ‘NONSYNDROMIC’ AUTISM

An important distinction in the genetics of autism is that between ‘syndromic’ and ‘non-syndromic’ (or idiopathic) autism (Fig. 2). ‘Non-syndromic autism’ is a term used to describe cases where autism is the primary diagnosis and is caused by unknown genetic or environmental cause, oligogenic, polygenic, and multifactorial mechanisms. Moreover, many genes involved in nonsyndromic intellectual disabilities (ID) and in epilepsy have also been implicated in the etiology of nonsyndromic ASD. These genes probably belong to a continuum of neurodevelopment disorders that manifest in different manners depending on associated genetic and environmental factors.

The term syndromic or secondary autism is used to refer to a condition caused by a well-known genetic variant, such as tuberous sclerosis, Rett syndrome, fragile X syndrome or other medical genetic conditions. It is typically associated with malformations and/or dysmorphic features and unlike ‘idiopathic’ ASD, it shows a different male: female sex ratio.

However, none of these etiologies is specific to autism because each of them encompasses a variable proportion of individuals with and without autism.

CAUSES OF ‘SYNDROMIC’ AUTISM

Currently a genetic cause can be identified in 20% to 25% of children with autism. The known genetic causes (Table 1) include cytogenetically visible chromosomal abnormalities (~5%), copy number variants (CNVs) (i.e. submicroscopic deletions and duplications) (10-20%) and single gene disorders in which neurologic findings are associated with ASD (~5%).

Figure 1. The puzzling etiology of ASD.
CHROMOSOME ABERRATION AS A CAUSE

Cytogenetic changes are consistently reported as one of the most common identifiable causes of autism. Cytogenetic abnormalities visible with karyotype analysis are found in approximately 5% of children with ASD and another 3-5% can be identified by fluorescence in situ hybridization (FISH) techniques.21 The most commonly reported (recurrent) cytogenetic abnormalities found in persons with autism are the 15q11–q13 duplication (of the maternal allele) of the Prader-Willi/Angelman syndrome region (1-3%), deletions involving 7q, 22q13, 2q37, 18q, Xp and sex chromosome aneuploidies (47, XYY and 45, X / 46, XY mosaicism).25

COPY NUMBER VARIATIONS (CNV) AS A CAUSE

Array comparative genomic hybridization (aCGH) is steadily replacing high-resolution chromosome analysis and FISH in the evaluation of children with autism. Array CGH is designed to test for known deletion/duplication syndromes on the entire genome plus assessment of subtelomeric regions.21 Rare de novo and some inherited CNVs can contribute to the genetic vulnerability to ASD in as many as 10% of examined cases. CNVs can involve a single gene and act much as a sequence-level mutation or they can encompass several genes as part of a genomic disorder.24 Some CNVs associated with autism include 1q21 deletion, 2p15-2p16.1 deletion, 15q13 deletion/duplication, 16p11.2 deletion/duplication.23

Weiss et al.26 reported 16p11.2 deletions or duplications in approximately 1% of individuals with autism and 1.5% of children with developmental or language delays. The 16p11.2 deletion often occurs de novo, but may be transmitted from parent to child in an autosomal dominant manner. Of note, however, is that the same 16p11.2 CNVs can be observed in a variety of other disorders including schizophrenia, bipolar disorder, seizures, ADHD, and dyslexia, as well as apparently unaffected family members; thus, interpretation of the significance of this CNV can be difficult.21

Overall, CNVs are linked to a broad variety of clinical features, including major or minor malformations, facial dysmorphisms, severe neurological symptoms, full-blown autism, milder autism-spectrum traits, or even behavioral disorders outside of the autism spectrum. Thus, the variable penetrance and great phenotypic heterogeneity characterizing CNV expressivity make it often difficult to determine whether in a given patient a CNV is the sole cause of autism, confers vulnerability to the disease, or represents a chance finding. The majority of CNVs are inherited from either one of the parents, who may display some autistic traits, but clearly without satisfying the criteria for autistic disorder. Notably, many CNVs found in ASD patients have been found also in patients with other psychopathologies, especially intellectual disability and schizophrenia.23

Figure 2. Causes of ‘syndromic’ and ‘non-syndromic’ autism.
### Table 1. Some genetic syndromes associated with autism

<table>
<thead>
<tr>
<th>Specific genetic disorder</th>
<th>Gene(s)/chromosome region involved</th>
<th>Autistic signs</th>
<th>Prevalence</th>
<th>Estimated rate (%) of autism in the disease</th>
<th>Estimated rate (%) in autism</th>
<th>Mental retardation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fragile X syndrome</td>
<td>FMR1</td>
<td>Poor eye contact, social anxiety, language impairment, stereotyped behaviors</td>
<td>1/3500 - 1/9000</td>
<td>18 - 33%</td>
<td>1-3%</td>
<td>Variable</td>
</tr>
<tr>
<td>Tuberous sclerosis</td>
<td>TSC1, TSC2</td>
<td>learning difficulties, behavioral problems</td>
<td>1 - 1.7/10000</td>
<td>25 - 60%</td>
<td>1 - 4% (8 - 14% if seizures are present)</td>
<td>Variable</td>
</tr>
<tr>
<td>Rett syndrome</td>
<td>MECP2</td>
<td>Stereotyped behaviors, language impairment, disturbance in social relatedness, loss of eye contact</td>
<td>1/8500 females</td>
<td>80 - 100%</td>
<td>&lt; 5%</td>
<td>Severe</td>
</tr>
<tr>
<td>Untreated phenylketonuria</td>
<td>PAH</td>
<td>Self-injurious behavior, lack of social responsiveness</td>
<td>1/10000 - 1/15000</td>
<td>-</td>
<td>5.70%</td>
<td>Severe</td>
</tr>
<tr>
<td>Prader-Willi syndrome</td>
<td>Del paternal allele at 15q11–q13</td>
<td>Repetitive behaviour and social deficits</td>
<td>1/10000 -1/30000</td>
<td>19 - 36.5%</td>
<td>1%-3%</td>
<td>Severe</td>
</tr>
<tr>
<td>Angelman syndrome</td>
<td>Del/mut in maternal UBE3A</td>
<td>Severe speech impairment stereotyped behaviors, immutability</td>
<td>1/10000 - 1/12000</td>
<td>50 - 81%</td>
<td>≤ 1%</td>
<td>Severe</td>
</tr>
<tr>
<td>Williams-Beuren syndrome</td>
<td>7q11.23 del</td>
<td>Social communication impairment, ranging from excessive talkativeness and overfriendliness to absence of verbal language and poor social relationships</td>
<td>1/7500 - 1/25000</td>
<td>7%</td>
<td>&lt; 1%</td>
<td>Variable</td>
</tr>
<tr>
<td>Smith-Magenis syndrome</td>
<td>17p11.2 del</td>
<td>Self-injurious behavior, stereotyped behaviors (self-hugging), immutability</td>
<td>1/15000</td>
<td>93%</td>
<td>&lt; 1%</td>
<td>Variable</td>
</tr>
<tr>
<td>Velocardiofacial/Di George syndrome</td>
<td>22q11.2 del</td>
<td>Speech delay, social skills difficulties</td>
<td>1/4000 - 1/6000</td>
<td>20 - 31%</td>
<td>&lt; 1%</td>
<td>Variable</td>
</tr>
<tr>
<td>Phelan-McDermid syndrome</td>
<td>22q13.3 del</td>
<td>Moderate to severe delays and often do not develop functional language.</td>
<td>unknown</td>
<td>50 - 70%</td>
<td>&lt; 1%</td>
<td>Severe</td>
</tr>
</tbody>
</table>
SINGLE GENE DISORDERS AS A CAUSE

Well-known genetic disorders can encompass autistic features in their clinical presentation, such as fragile X syndrome, tuberous sclerosis, neurofibromatosis, untreated phenylketonuria, and Rett syndrome. The most prevalent single gene disorders as a cause are fragile X syndrome (around 3-5%) and tuberous sclerosis (TSC1/TSC2 - around 1%).

Whereas 1% to 3% of children ascertained on the basis of an autism diagnosis have fragile X syndrome, at least half of the children with fragile X syndrome have some autistic behavior.

The situation is similar with tuberous sclerosis (TSC) - 25-50% of intellectually disabled individuals with TSC fulfill autism diagnostic criteria but only 1.1-1.3% of individuals initially diagnosed with ASD have TSC.

Notably, the clinical manifestations of 'syndromic' autism can be highly heterogeneous, even in the presence of the same well characterized mutation or genomic rearrangement, likely due to differences in genetic background and epigenetic influences.

THE SEARCH FOR CAUSE OF 'NON-SYNDROMIC' AUTISM

Many genes and environmental factors are likely to contribute to the etiology of autism, making it difficult to isolate disorder genes. The search methods of the disorder susceptibility genes include linkage studies, genome-wide association studies, whole-exome sequencing, expressions analysis and epigenetic studies.

LINKAGE STUDIES

Substantial effort in autism genetics over the last 10 years has been focused on genetic linkage analysis using an affected sibling-pair design in multiplex families. Linkage can be defined as the tendency for alleles close together on the same chromosome to be transmitted together, as an intact unit, through meiosis. Linkage studies are either performed as full genome screens with a dense set of genetic markers covering all chromosomes, or locally (fine-mapping) at a certain chromosomal area of interest.

To date, only loci on 17q11–17q21 and 7q22-7q32 have been replicated at levels that could be considered genome-wide significant. Both loci contain some of the most intensively studied candidate genes associated with ASD - CNTNAP2, RELN and MET.

The linkage data indicate that many loci may underlie risk of autism, which is consistent with the well-accepted hypothesis that many genes may be associated with autism. To address the limitation of reduced linkage signals due to marked genetic heterogeneity, some studies have made an effort to increase sample homogeneity by focusing on selected characteristics (or endophenotypes) of autism, such as language measures (e.g., age at first word), developmental milestones (e.g., bladder and bowel control), developmental regression, and restricted repetitive and stereotyped patterns of behavior, interests, and activities (e.g., obsessive-compulsive behavior). Such endophenotypes, which can also be found in nonautistic family members and in the general population, facilitate the mapping of quantitative trait loci (QTL) that individually contribute to the overall autism phenotype.

QTL studies suggest that 'subtyping' autism patients will help narrow the search for specific autism susceptibility genes. The fact that linkage studies have not yet identified 'the ASD gene' is not a failure of these methods but rather an accurate reflection of the complexity of this disorders and the need for larger homogenized sample sizes.

GENOME-WIDE ASSOCIATION STUDIES (GWAS)

GWAS studies identify single-nucleotide polymorphisms (SNPs) and other common genetic variants in DNA that may be associated with a disease or trait by investigating the entire genome using an unbiased hypothesis-free search. The first GWAS was carried out by Weiss et al. and they have found no statistically significant SNP association in over 1000 families. The authors then carry out a genotyping of the most significant SNPs in additional autistic families and a point mutation in 5p15 have shown statistical significance. This SNP fell between genes encoding SEMA5A (a member of the semaphorin family of proteins involved in axon guidance) and TAS2R1 (a bitter-taste receptor). Independent evidence was presented of reduced expression of SEMA5A in lymphoblastoid cell lines, whole-blood and brain samples of autistic individuals compared to control.

Wang et al. carried out another GWAS. They used two populations: one largely overlapping with the sample used by Weiss et al. and the other including over 1200 cases and nearly 6500 controls. One region of genome-wide significance was identified at 5p14. The association signal came...
from a region between CDH10 and CDH9, two genes encoding neuronal cell-adhesion molecules.\textsuperscript{33} In 2010 the Autism Genome Project (AGP) Consortium genotyped 1 million SNPs and analyzed 1558 defined ASD families\textsuperscript{34}, identifying genome-wide association with rs4141463, located in the MACROD2 gene, as crossing a preset significance threshold of $P < 5 \times 10^{-8}$.

The different results gathered from the independent GWAS, despite the great wealth of data gained from each single study, have been interpreted as stemming from the large phenotypic and genetic heterogeneity of ASD. Thus, current efforts are aimed at finding autism subtype-related genetic variants, also by searching for QTL. Nevertheless, GWA studies may help identify targets for more in-depth studies of etiological mechanisms.

WHOLE-EXOME SEQUENCING (WAS)

Since 2005, next-generation sequencing (NGS) technologies have been improving as rapid, high-throughput and cost-effective approaches to fulfill medical sciences and research demands.\textsuperscript{35} Whole-exome sequencing (WES) has recently been introduced to identify rare or novel genetic defects from genetic disorders. Particularly, ASD is a model disease to apply WES because multiple loci are involved in its development with relatively weak genotype-phenotype correlation.\textsuperscript{23} The data gathered from WES\textsuperscript{35-38} demonstrates the importance of de novo mutations in the etiology of ASD. There may be several hundred genes in which high-risk-conferring de novo mutations can occur and the vast majority of them may increase the risk but does not “cause” the disease, further supporting an oligogenic/polygenic model. Many of the disrupted genes were found to impact important gene networks (synaptic plasticity, -catenin/chromatin remodeling), and several de novo mutations were found in genes previously implicated in other neurodevelopmental disorders and in intellectual disability (e.g., SCN1A, SCN2A, GRIN2B).\textsuperscript{35-38}

Other use of WAS is to find candidate recessive mutations in autistic probands with known shared parental ancestry.\textsuperscript{23} Chahrour et al.\textsuperscript{39} used homozygosity analysis to identify probands from non-consanguineous families that showed evidence of distant shared ancestry, suggesting potentially recessive mutations. They identified four novel candidate genes, namely UBE3B, CLTCL1, NCK-AP5L and ZNF18, which encode proteins involved in proteolysis, GTPase-mediated signaling, and cytoskeletal organization.\textsuperscript{39} Puffenberger et al.\textsuperscript{40} found a homozygous missense mutation in HERC2 (c.1781C>T, p.Pro594Leu) associated with global developmental delay and ASD. Novarino et al.\textsuperscript{41} screening families with autism, epilepsy, and intellectual disability identified homozygous inactivating mutations in the BCKDK gene ( Branched Chain Ketoacid Dehydrogenase Kinase).

The data acquired by WES shows that lots of genes have been identified and associated with ASD but no common etiology can be proposed. All aspects such as oligogenic, de novo mutations, recessive inherited mutations, polygenic, multifactorial, pleiotropic effects, combination of locus heterogeneity should be considered when thinking of ASD etiology.

GENOME WIDE EXPRESSION STUDIES

The advent of microarray technologies allows the design and implementation of genome-wide expression profiling which helps to unravel the molecular basis of phenotypic variation in many disease states. Global expression profiling provides the opportunity to evaluate possible transcriptomic deviations in ASD at both genes and gene-network levels. Currently, several published studies describe genome-wide expression patterns in biomaterials derived from ASD patients and controls.\textsuperscript{42} So far, only a handful of studies have investigated the ASD transcriptome in post-mortem brain samples\textsuperscript{43,44} and other studies have used RNAs from lymphoblastoid cell lines or blood\textsuperscript{45-50} demonstrating their usefulness for this purpose. Lymphoblasts provide access to larger sample sizes and unlimited source of RNA. Moreover, for technical reasons lymphoblasts can be better controlled for various experimental confounds that may influence the fidelity of measuring gene expression data.\textsuperscript{51}

Most of the studies found different expression in genes that are involved in the neurodevelopment in ASD samples compared to controls. NAGLU, FLAP/AOX5AP, CHL1, and ROBO1 that are involved in neuronal differentiation, brain development and axon guidance were found downregulated by Hu et al.\textsuperscript{47} GFG12, MYT1L, and GAS7 that are associated with neuronal differentiation and outgrowth were found downregulated by Garbett et al.\textsuperscript{42} Semaphorin 4C (SEMA4C) and glutamate dehydrogenase (GLUD1) that are involved in neural development and in glutamatergic neurotransmission were discovered upregulated by Gregg et al.\textsuperscript{46}

Several of these studies\textsuperscript{43,46,47,49,52} also found different expression in the genes involved in the immune system: pro-inflammatory cytokines, NK
and CD8+ cell-related genes including receptor genes and genes involved in the cytotoxicity pathways - ASS, DAPK1, FLT1, IL6ST, IL2RB, TH11TH2, FAS, SPON2, IL2RB, PRF1, EAT2/SH2D1B, GZMB, CX3CR1 PAM, suggesting neural inflammation model for ASD.

The data acquired so far suggest that genes involved in the nervous system development and the immune system may play a role in the pathogenesis of ASD. Analysis of peripheral tissues such as blood and lymphoblast cell lines are important as a complement to SNP genotyping and CNV analyses in order to assess the functional significance of genetic variants, and identify ASD biomarkers.

**EPIGENETIC STUDIES**

Epigenetic mechanisms, such as DNA methylation, modifications of histone proteins and miRNA, regulate chromatin structure and/or gene expression without changing DNA sequence. Epigenetic abnormalities are associated with several neurodevelopmental diseases including ASD. Regulation of neuronal structure and function through epigenetic mechanisms is believed to be critical in the development of the nervous system.

Regulation of gene expression through DNA methylation has been found in several genes associated with ASD. Hyper-methylation of specific CpG sites in the promoter regions of BCL-2 and RORA that leads to downregulation of these genes was identified in autistic children compared to healthy developing twins. SHANK3 is another gene that is associated with ASD and has an epigenetic regulation.

Alternation in the modifications of the histone proteins is another epigenetic mechanism that could be related to ASD. SMCX gene, which encodes histone 3 lysine 4 (H3K4) me3 - specific demethylase, regulates through demethylation of the histones other genes, i.e., SCN2A, CACNA1H, BDNF, SLC18A1, associated with ASD and cognitive dysfunction.

Another aspect of epigenetic control of gene expression is through miRNA. There are only a handful of studies that have analyzed the expression profile of miRNA in ASD samples compared to healthy controls. They all have found differently expressed miRNA and their predicted target genes in the ASD samples suggesting that miRNA may play a role in the etiology of this disorder.

The role of epigenetics in ASD has only emerged in the last few years, and represents a growing area of research.

**CONCLUSION**

The latest advances in the discovery of the genetic factors involved in the etiology of ASD have increased significantly over the past years. A genetic factor in the etiology of this disorder is certain, as certain as its complexity is. Understanding the genetic factors involved is crucial to establishing future intervention strategies. It is expected that high-throughput molecular screenings, such as high resolution array-CGH, whole-exome and whole-genome sequencing, as well as transcriptomic and epigenetic analysis, will further increase our understanding of the genetics of ASD. The interpretation of this newly acquired genetic data will require multidisciplinary approach, but still clinical geneticists have an important role in the diagnosis and research of autism.

**REFERENCES**