SMGroup

Post-genomic Era in Autistic Spectrum Disorder

Oh S[1], Lin PI[2], Barzman D[3], Ji H[4] and Hutton J[5]*

[1]Department of Computer Science and Statistics, Jeju National University, S Korea.

[2]Division of Biostatistics and Epidemiology, Department of Pediatrics, Cincinnati Children's Hospital Medical Center, Cincinnati, USA

[3]Division of Psychiatry, Department of Pediatrics, Cincinnati Children's Hospital Medical Center, Cincinnati, USA

[4]Division of Asthma Research, Director of pyrosequencing core, Department of Pediatrics, Cincinnati Children's Hospital Medical Center, Cincinnati, USA

[5]Division of Biomedical Informatics, Department of Pediatrics, Cincinnati Children's Hospital Medical Center, Cincinnati, USA

***Corresponding author:** John Hutton, Division of Biomedical Informatics, Department of Pediatrics, Cincinnati Children's Hospital Medical Center, Cincinnati, USA, Tel: 513-636-0265; FAX : 513-636-2056; E-mail: john.hutton@cchmc.org

ABSTRACT

Despite the findings on the genetic basis for autism spectrum disorder, there have been only a handful of variants that have been consistently found to confer the risk of autism spectrum disorder, and these variants are found in only a small fraction of patients. The difficulty of unraveling the genetic architecture underlying the etiologies of autism spectrum disorder may stem from the heterogeneity of phenotypes in patients, the complexity of genetic variants, and epigenetic phenomena that jointly contribute to its etiologies. Over the last decade, many genomic technologies have been employed to uncover the underlying biological and molecular mechanisms underlying autism spectrum disorder in the context of diagnoses and therapeutic

pathophysiology. Nevertheless, it still remains elusive to provide generalized gold-standard specific biomarkers for autism spectrum disorder that explain the range of phenotypes and the specific symptoms in an individual patient. This chapter reviews genomic studies and their influences on our concepts of the biology of the disorder.

Keywords: Autism spectrum disorder; ASD; Genomics; Biomarkers

ABBREVIATIONS: ASD: Autism Spectrum Disorder; **DSM-5:** Diagnostic and Statistical Manual of Mental Disorders; **SNP:** Single Nucleotide Polymorphism; **CNV:** Copy Number Variation; **GWAS:** Genome-Wide Association Study; **WES:** Whole Exome Sequencing; **fMRI:** Functional Magnetic Resonance Imaging; **hiPSC:** human induced pluripotent stem cell

INTRODUCTION

According to DSM-5, individuals with autism spectrum disorder (ASD must show symptoms from early childhood. ASD is a complex neurodevelopmental disorder of the brain that can cause problems with thinking, feeling, language, and the ability to relate to others. Hallmark characteristics are impairment of social communication and restricted, repetitive patterns of behavior. Thus far, a number of groups have identified rare and common variants among genes and their effect on risk of autism [1-15], including SNPs, chromosomal abnormalities, CNVs, alternative splicing and allelic specific expression, disruption of target genes, protein interactions, and their functional effects. More recently, owing to the rapid growth of technologies, whole-genome-wide and whole-exome sequencing have been successfully conducted based on family and twin studies in order to detect rare/common inherited variants, and *de novo* mutations that are directly associated with the molecular changes associated with pathogenic complexity of ASDs [1,10,16-20]. In addition, meta-framed strategies such as integration between transcriptional profiling and targeted exome sequencing or GWASs have given rise to better characterization of the cascades of susceptible genes and disease risk [9,14,21-23]. The studies of epigenome in autistic children have also provided possible mechanism(s) accounting for susceptibility and heritability of ASD (recently reviewed [24]). There is a gradual increment of pervasiveness in prevalence of disease in the US population (and also, over the world) and the conclusions of many studies are not well-replicated between studies and do not provide complete answers for personalized genome data. In addition, establishment of study protocols and rigorous analytical frameworks for ASD studies are currently necessitated in the field as it is in its infancy. Collectively, we outline an influential role of genomic strategies in ASD clinical studies, currently issued limitations, and how integrated meta-genomic tools have emerged as new paradigm and future directions to go ahead in the upcoming years.

BIOMARKERS OF AUTISM SPECTRUM DISORDER

Genome-wide Association Studies (GWAS)

There are two major genetic-based approaches to identify abnormally formed mutations between genes, composing of rare de novo and heritably transmitted mutations from paternal and maternal effects of origin [25] and common variants. Basically, there are two types of studies. Two or more affected autistic children in a family can be studied along with parents or an affected child in given probands from ASDs patients can be compared to normal control individuals not in families. For the purpose of examination of such variants disrupting targeted protein coding genes and non-coding regions, whole-genome wide association studies (GWASs) have been addressing hypothetical questions on mutations at specific loci that are much fewer, but are more commonly detected markers than rare variants in autistic cases by scanning whole-genome sequences profiles [1,2,10,13,19,26-29]. It aims at the characterization of interplay between single-nucleotide polymorphisms and varying ASD phenotypes in patients by examining particular candidate SNPs that differ in allele frequency at genetic loci between cases and controls to be associated with detrimental effects on ASD risk. Strikingly, 5p14.1 between gene CDH9 and CDH10 and 5p15.2 between SEMA5A and 5AS2R are known autistic specific locus detected by statistical testing in GWAS [10]. The approach further enables a determination of the extent to which phenotype homogeneity is represented by common variants as the criteria feature of intra-ASD classes, subsequently grouped in other complementary factors such as verbal IQ.

There are ~1,000 ASD-specific genes related to the causal and consequent mal-functionalities, relying upon mainly different genetic background and secondarily affected either independently or accumulatively by epigenetic and/or environmental risk factors [6,21,23,30,31]. It is well known that common variants have been more specifically supportive over a wide range of phenotypic spectrum variance.

Although GWAS studies set out to identify variants that contribute to the risk of ASD, the replication of findings has been difficult because of sample size, types of variants (e.g., common versus rare variants), sources of samples (e.g., case-control versus family-based populations), gene-gene, and gene-environment interactions, and epigenomic phenomena. The inconsistent results across different studies have become a bottleneck for meta-analysis strategies that aims to increase the statistical power of detection of inherited common variants that are associated with ASD.

In this perspective, current GWASs give us a great opportunity to further fine-tune signal-to-noise in terms of measuring more reliably genotyped allelic mutations and systematic validation procedures for entire protocols on such mutated variations with well-annotated functional impacts. Also, importantly, non-coding variants that have played a key role for characterization of disease mechanisms have been emerged as the next stage to explore the extended regions for active promoters and strong enhancers as well as coding variants.

Whole Exome Sequencing

Whole-exome techniques based on ultra-high throughput sequencing methods have been capturing highly promising subset of genome that encode protein-coding regions by means of diagnostic testing and clinical therapy on complex disease with inherent traits such as autism [4,16,20,21,26,27,32]. Despite the great value and promise of exome sequencing, identified known and novel findings for *de novo* and inherited rare variants have been unable to provide a complete series of causes for ASD-specific genomic variation.

To date, it has been widely reported that *de novo* loss-of-function rare variants that are relevantly mutated in autistic patients are generally involved with underlying pathophysiology, chromatin remodeling pathway, synaptic functions, and transcriptional processes. Far more rare SNPs or CNVs with stronger evidence for disease risk are detected in a lesser extent in the autistic cases. Since autism is one of highly inherited neurodevelopmental and neuropsychiatric disorders, it implies that *de novo* loss-of-function (transmitted or un-transmitted from unaffected parents) coupling with characterization of inherited variants will be more valuable to provide better insights for genetic causes of ASD in the format of an aggregated disease effect.

To this end, it is pivotal to snapshot more comprehensive results revealing information on causes of ASDs by *de novo* loss-of function, missense transmitted or non-transmitted mutations from parents of origin through sophisticated strategies. This will enable meta-framed analyses for the diagnostic testing and clinical treatments in a wide spectrum of genotypic and phenotypic autistic patients.

Accordingly, first of all, more improved and enhanced bioinformatics and statistical methods are immediately needed to precisely measure and test rare and common variants so as to avoid misleading conclusions and to reduce false discoveries.

Secondly, although there is a huge list of disease-causative candidate genes, most genes are limited to a particular family-unique outcome and the focus on causes of autism risk could be data-driven biased results. At this point, intra-platform and inter-platform meta approaches are more appealing to derive more comprehensive landscape of relationship between aberrant mutations and resultant phenotypes, detailed characterization of varying phenotypic variance over the spectrum, led by various genetic, epigenetic, and environmental factors. Examples of these approaches include combinatory strategies between multiple GWAS studies to increment the power of detection of autistic specific markers and validate repeatedly induced abnormal functionalities between studies, or between transcriptome and exome-sequencing/genome-sequencing.

Thirdly, due to the enormous difficulties of collecting DNA samples from primary brain tissues in autistic patients, it mostly depends on post mortem or lymphoblast cell lines. More recently, as alternative, human induced pluripotent stem cell (hiPSC) approaches have been attractively

proposed by directly targeting patient-specific neuronal cells induced *in vitro* [33-36]. Patient-specific iPSCs can be generated from tissues easily accessible from patients such as fibroblast and blood cells using standard methods, and then re-differentiated into different subtypes of neurons [37]. This approach recapitulates human neurodevelopment in a personalized manner in vitro, which facilitates the understanding of patient-specific autism-associated genomic variations on neurodevelopment and on autism pathogenesis [38]. In combination with emerging genetic editing tools such as sequence-specific designed zinc finger nuclease (ZFN), transcription activator-like effector nuclease (TALENs), or CRISPR/Cas-9, the iPSC models can explore the promise of gene and epigenome correction for therapeutic purposes. In addition, fibroblasts and blood cells can be converted to neuronal cells by expressing defined transcriptional factors [39], although standardization of protocols to stabilize these cells are needed. Once established, the iPSC derived autistic patient-specific neuronal cells can be used for early stage drug discovery, including quantitative biochemistry, functional genomics, proteomics, and high-throughput and high-content chemical screening [35]. Although many challenges remain, iPSC-derived neuronal cells hold great promise for modeling autism and drug discovery, when integrated with clinical disease pathophysiology.

Transcriptome

Transcriptome research sheds more light on how genomic variations contributes to phenotypic variations. The uncertainties in the impact of DNA variants associated with variable isoforms, trans-/cis-regulatory, or epigenetic effects, often need to be resolved with a deeper understanding of the transcriptome. However, just like most neuropsychiatric disorders with a pediatric onset, gene expression studies on ASD are faced with challenges, such as sources of tissues and timing of mRNA analysis, to name a few of them. Several investigators have attempted to justify the value of mRNA levels in peripheral blood cells in in ASD [32,33]. However, the concerns about inferring the genomic functions associated with brain phenotypes using peripheral blood cells have been highlighted by at least two lines of research. First, some genes are expressed substantially more in human brains, compared to mouse brains [34] and other primate brains [35,36] – of which the expression patterns are not observed in other tissues. Second, only 23% of gene expressions in the genome in peripheral blood cells are correlations with those in the brain [37], while some highly heritable variants may cause correlated gene expressions between brain and spleen [38]. Taken together, these lines of evidence have cautioned that transcriptomic research based on peripheral tissues may yield limited information on genomic mechanism underlying ASD. Furthermore, gene expression may also differ across different brain regions; such a variation results from structural and functional differences [39]. Therefore, gene expressions in brain regions with well-validated neuroimaging findings associated with the ASD-related phenotypes should have more convergent evidence for their roles in etiologies of ASD. For example, fMRI studies of ASD have demonstrated dysfunctional activation in brain regions of restricted and repetitive behaviors (RRBs) and social communication [40]. One of the most replicated findings has been dysfunction of social

brain regions. This dysfunction may be due to poor preference for social stimuli rather than a primary dysfunction in social brain regions [41]. These neuroimaging findings should serve as endophenotypes more closely associated with transcriptome, compared to the diagnosis of ASD.

In addition to sources of mRNA and brain-region-specific expressions, transcriptome studies across different populations may generate inconsistent results, since gene expressions can be impacted by various conditions, such as co-occurring physiological conditions and environmental factors. Despite these challenges, several studies investigating genes expressed in brain tissues have revealed some novel findings. For example, Gupta and colleagues reported that some neuronal and microglial genes were dysregulated in autistic cortical brain based on their RNA sequencing results [42]. Interestingly, a meta-analysis showed that gene expression studies on brain tissues seem to be more likely to generate replicable findings that studies using blood samples in ASD [43]. These findings further highlight the importance of examining gene expressions in brain tissues, rather than other peripheral tissues.

DISCUSSION AND CLOSING REMARKS

Despite efforts of investigators in the field, there does not exist a method or strategy to define gold standard biomarkers for phenotypes of patients with autistic spectrum disorders. Limitations of these efforts include, for example: insufficient sample size to derive a solid conclusion, failure to replicate outcomes between similar studies, population and patient-specific conclusions, difficulties to collect corresponding brain tissues. In addition, many studies lacked focus on comprehensive repertoires including cis-regulatory elements, post-transcriptional and epigenetic modifications, and platform-by-platform meta approaches. In order to overcome these limitations and better understand pathophysiology of biomarkers in autism spectrum disorders, a few groups have been accelerating the ways to define functional consequences of differential expression of genes that are associated with the dysregulation in autism. Followed by either transcriptome or exome sequencing approaches for the detection of the functional effects, more enhanced analytical strategies at higher levels such as post-transcriptional modifications in order to better understand tissue-specific (disease-specific) pathologies, allelic specific expression via genomic imprinting and cis-acting regulation, enables investigators to unravel pathway/network modules by focusing on the allelic imbalance of rare deleterious variants between autistic versus normal samples. Likewise, the ample knowledge of alternatively spliced variants that produce distinct protein products and corresponding functionalities could be incorporated with genomic data, such as whole genome wide association studies and whole exome sequencing approaches. Differential splicing across tissues elucidates the link of genotype-phenotype patterns, under the given hypothetical assumption that deregulated patterns of splicing variants could be related to neuronal dysfunctional mechanisms such as ASD-specific splicing markers, A2BP1/FOX1.

Taken together, transcriptional level approaches are further confirmed with disease-causing genes explored in GWAS or whole exome sequencing in an integrated manner. Beyond gene levels,

alternative splicing and allelic imbalance as well as intronic regions containing cis-regulatory elements that could not be captured by targeted exome sequencing are further characterized to find the common candidate and putative disease markers at the higher system biology level. While meta-framed system biology level strategy has more variability and divergent variance across distinct tissues in pathogenic biomarkers of autism, it is a more comprehensive and systematic method to efficiently define the underlying molecular mechanisms when compared to simple gene level transcriptome approach or gene disruption and mutation based analysis in independent genomic studies.

At this point, the primary merit of integrated genomic approaches with transcriptomic profiles is to strengthen the assembly of various scenarios on the functional impacts for the altered expression because of, for example, structural variants, copy number variations, allelic specific expression across distinct types of tissues/cells/developmental stages/disease progressive time points, post-transcriptional procedures on the basis of alternative splicing and RNA-editing, loss-of-function rare deleterious variants. Curated catalogues of gold standard biomarkers of autistic spectrum disorders should account for genomic, epigenomic, and environmental risk factors to draw a complete landscape of primary and secondary phenomena as well as aggregated effects for causes of disease risk. In the coming years, large-volumes of data will need to be well-annotated and classified for either explicitly different or likely to be similar phenotypes within the spectrum of autism accounting for the inherent features from patient to patient. This high quality repository of candidate disease-causing genes and variants will support stringent and robust computational and statistical methodological approaches to identify causes of specific types of autism within the autism spectrum.

References

1. Anney R, Klei L, Pinto D, Regan R, Conroy J. A genome-wide scan for common alleles affecting risk for autism. Hum Mol Genet. 2010; 19: 4072-4082.

2. Cai G, Edelmann L, Goldsmith JE, Cohen N, Nakamine A, Reichert JG, et al. Multiplex ligation-dependent probe amplification for genetic screening in autism spectrum disorders: efficient identification of known microduplications and identification of a novel microduplication in ASMT. BMC Medical Genomics. 2008; 1: 50.

3. De Rubeis S, He X, Goldberg AP, Poultney CS, Samocha K. Synaptic, transcriptional and chromatin genes disrupted in autism. Nature. 2014; 515: 209-215.

4. Griswold AJ, Dueker ND, Van Booven D, Rantus JA, Jaworski JM, Slifer SH, et al. Targeted massively parallel sequencing of autism spectrum disorder-associated genes in a case control cohort reveals rare loss-of-function risk variants. Mol Autism. 2015; 6: 43.

5. Hu J, Liao J, Sathanoori M, Kochmar S, Sebastian J. CNTN6 copy number variations in 14 patients: a possible candidate gene for neurodevelopmental and neuropsychiatric disorders. J Neurodev Disord. 2015; 7: 26.

6. Koufaris C, Sismani C. Modulation of the genome and epigenome of individuals susceptible to autism by environmental risk factors. Int J Mol Sci. 2015; 16: 8699-8718.

7. Kukurba KR, Zhang R, Li X, Smith KS, Knowles DA. Allelic expression of deleterious protein-coding variants across human tissues. PLoS Genet. 2014; 10: e1004304.

8. Lal D, Ruppert AK, Trucks H, Schulz H, de Kovel CG, Kasteleijn-Nolst Trenité D, et al. Burden analysis of rare microdeletions suggests a strong impact of neurodevelopmental genes in genetic generalised epilepsies. PLoS Genetics. 2015; 11: e1005226.

9. Luo R, Sanders SJ, Tian Y, Voineagu I, Huang N. Genome-wide transcriptome profiling reveals the functional impact of rare de novo and recurrent CNVs in autism spectrum disorders. Am J Hum Genet. 2012; 91: 38-55.

10. Ma D, Salyakina D, Jaworski JM, Konidari I, Whitehead PL. A genome-wide association study of autism reveals a common novel risk locus at 5p14.1. Ann Hum Genet. 2009; 73: 263-273.

11. Song Y, Hakoda Y, Sanefuji W, Cheng C. Can They See It? The Functional Field of View Is Narrower in Individuals with Autism Spectrum Disorder. PLoS One. 2015; 10: e0133237.

12. Stamova BS, Tian Y, Nordahl CW, Shen MD, Rogers S. Evidence for differential alternative splicing in blood of young boys with autism spectrum disorders. Mol Autism. 2013; 4: 30.

13. Stepniak B, Kästner A, Poggi G, Mitjans M, Begemann M. Accumulated common variants in the broader fragile X gene family modulate autistic phenotypes. EMBO Mol Med. 2015; 7: 1565-1579.

14. Voineagu I, Wang X, Johnston P, Lowe JK, Tian Y. Transcriptomic analysis of autistic brain reveals convergent molecular pathology. Nature. 2011; 474: 380-384.

15. Wang L, Li J, Ruan Y, Lu T, Liu C. Sequencing ASMT identifies rare mutations in Chinese Han patients with autism. PLoS One. 2013; 8: e53727.

16. Bamshad MJ, Ng SB, Bigham AW, Tabor HK, Emond MJ. Exome sequencing as a tool for Mendelian disease gene discovery. Nat Rev Genet. 2011; 12: 745-755.

17. Hussman JP, Chung RH, Griswold AJ, Jaworski JM, Salyakina D. A noise-reduction GWAS analysis implicates altered regulation of neurite outgrowth and guidance in autism. Mol Autism. 2011; 2: 1.

18. Metzker ML. Sequencing technologies - the next generation. Nat Rev Genet. 2010; 11: 31-46.

19. Shi L, Zhang X, Golhar R, Otieno FG, He M. Whole-genome sequencing in an autism multiplex family. Mol Autism. 2013; 4: 8.

20. Toma C, Torrico B, Hervas A, Valdes-Mas R, Tristan-Noguero A, Padillo V, et al. Exome sequencing in multiplex autism families suggests a major role for heterozygous truncating mutations. Molecular Psychiatry. 2014; 19: 784-790.

21. Codina-Sola M, Rodriguez-Santiago B, Homs A, Santoyo J, Rigau M, Aznar-Laín G, et al. Integrated analysis of whole-exome sequencing and transcriptome profiling in males with autism spectrum disorders. Molecular Autism. 2015; 6: 21.

22. Zhao Q, Kirkness EF, Caballero OL, Galante PA, Parmigiani RB, Edsall L, et al. Systematic detection of putative tumor suppressor genes through the combined use of exome and transcriptome sequencing. Genome Biology. 2010; 11: R114.

23. Daimon CM, Jasien JM, Wood WH 3rd, Zhang Y, Becker KG. Hippocampal Transcriptomic and Proteomic Alterations in the BTBR Mouse Model of Autism Spectrum Disorder. Front Physiol. 2015; 6: 324.

24. Oh S, Ji H, Barzman D, Lin PI, Hutton J. Pediatric asthma and autism-genomic perspectives. Clin Transl Med. 2015; 4: 37.

25. Migliavacca E, Golzio C2, Männik K, Blumenthal I, Oh EC. A Potential Contributory Role for Ciliary Dysfunction in the 16p11.2 600 kb BP4-BP5 Pathology. Am J Hum Genet. 2015; 96: 784-796.

26. Adzhubei I, Jordan DM, Sunyaev SR. Predicting functional effect of human missense mutations using PolyPhen-2. Curr Protoc Hum Genet. 2013; Chapter 7: Unit7.

27. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A. A method and server for predicting damaging missense mutations. Nat Methods. 2010; 7: 248-249.

28. MacArthur DG, Balasubramanian S, Frankish A, Huang N, Morris J. A systematic survey of loss-of-function variants in human protein-coding genes. Science. 2012; 335: 823-828.

29. Warrier V, Chakrabarti B, Murphy L, Chan A, Craig I. A Pooled Genome-Wide Association Study of Asperger Syndrome. PLoS One. 2015; 10: e0131202.

30. Melke J, Goubran Botros H, Chaste P, Betancur C, Nygren G. Abnormal melatonin synthesis in autism spectrum disorders. Mol Psychiatry. 2008; 13: 90-98.

31. Zeng L, Zhang Q, Li S, Plotnikov AN, Walsh MJ. Mechanism and regulation of acetylated histone binding by the tandem PHD finger of DPF3b. Nature. 2010; 466: 258-262.

32. Pritchard JK. Are rare variants responsible for susceptibility to complex diseases? Am J Hum Genet. 2001; 69: 124-137.

33. Acab A, Muotri AR. The Use of Induced Pluripotent Stem Cell Technology to Advance Autism Research and Treatment. Neurotherapeutics. 2015; 12: 534-545.

34. Lim CS, Yang JE, Lee YK, Lee K, Lee JA. Understanding the molecular basis of autism in a dish using hiPSCs-derived neurons from ASD patients. Mol Brain. 2015; 8: 57.

35. Haggarty SJ, Silva MC, Cross A, Brandon NJ, Perlis RH. Advancing drug discovery for neuropsychiatric disorders using patient-specific stem cell models. Mol Cell Neurosci. 2016;.

36. Mohan KN. Stem Cell Models to Investigate the Role of DNA Methylation Machinery in Development of Neuropsychiatric Disorders. Stem Cells Int. 2016; 2016: 4379425.

37. Kim DS, Ross PJ, Zaslavsky K, Ellis J. Optimizing neuronal differentiation from induced pluripotent stem cells to model ASD. Front Cell Neurosci. 2014; 8: 109.

38. Wang P, Lin M, Pedrosa E, Hrabovsky A, Zhang Z. CRISPR/Cas9-mediated heterozygous knockout of the autism gene CHD8 and characterization of its transcriptional networks in neurodevelopment. Mol Autism. 2015; 6: 55.

39. Xu J, Du Y, Deng H. Direct lineage reprogramming: strategies, mechanisms, and applications. Cell Stem Cell. 2015; 16: 119-134.

40. Ha S, Sohn IJ, Kim N, Sim HJ, Cheon KA. Characteristics of Brains in Autism Spectrum Disorder: Structure, Function and Connectivity across the Lifespan. Exp Neurobiol. 2015; 24: 273-284.

41. Philip RC, Dauvermann MR, Whalley HC, Baynham K, Lawrie SM. A systematic review and meta-analysis of the fMRI investigation of autism spectrum disorders. Neurosci Biobehav Rev. 2012; 36: 901-942.

42. Gupta S, Ellis SE, Ashar FN, Moes A, Bader JS, et al. Transcriptome analysis reveals dysregulation of innate immune response genes and neuronal activity-dependent genes in autism. Nat Commun. 2014; 5: 5748.

43. Ch'ng C, Kwok W, Rogic S. Meta-Analysis of Gene Expression in Autism Spectrum Disorder. Autism Res. 2015; 8: 593-608.