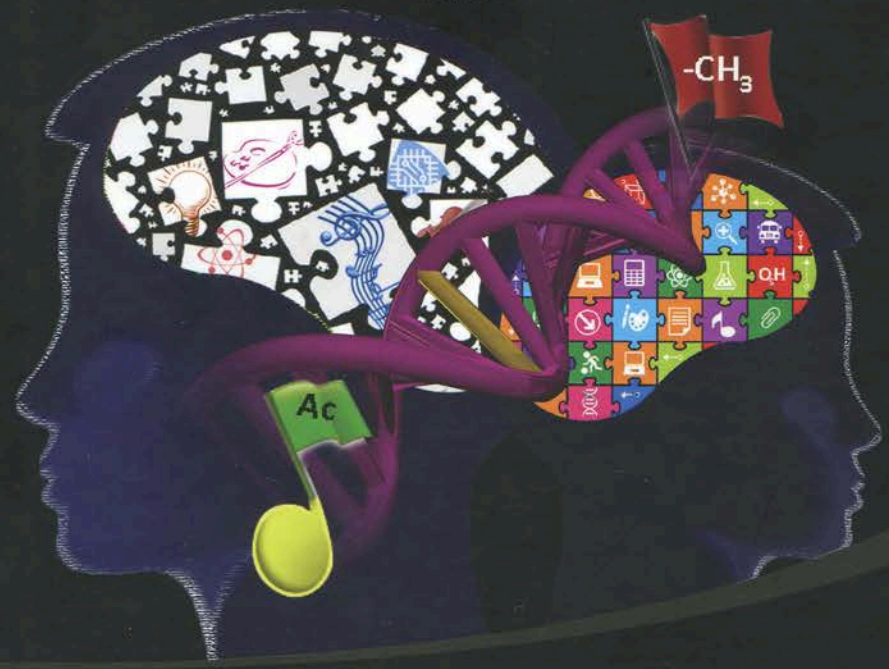
FRONTIERS IN AUTISM RESEARCH

New Horizons for Diagnosis and Treatment

Valerie W Hu

Editor



World Scientific

Chapter 5

The Impact of Integrative Unconventional Data Analysis Approaches on Advancing Autism Genetics Research

Zohreh Talebizadeh* and Ayten Shah

Abstract

The progress made in genomics technology provides unprecedented opportunity for the research community to accelerate discovering the genetic etiology of autism. A clear trend in the literature from gene-centric studies to whole-genome approaches is the best illustration of the incorporation of the state-of-the art techniques in autism research, which has progressed from the era of linkage scans, single gene and genome-wide association studies, and copy number variations to next-generation sequencing of the entire human genome. However, with the exception of studies that identified a few causative genes and a limited number of replicated findings, the rest have mainly contributed to the ever growing list of potential autism candidate genes, awaiting replication. This unexpected outcome has now become an additional challenge in moving forward from gene discoveries to diagnostic tests and treatment options in autism. In this chapter, we highlight findings derived from a few integrative approaches using unconventional models, which have resulted in promising discoveries and provided practical examples of how uncovering genetic causes of autism may be accelerated.

*Correspondence to: Children's Mercy Hospital, Kansas City, MO, USA
University of Missouri-Kansas City; School of Medicine; Kansas City; MO; USA
Email: ztalebi@cmh.edu

Keywords: Autism; Integrative approaches; Genetics; Epigenetics; Gene regulatory processes; Gender disparity; Phenotypic stratification; Systems biology; Interactomes.

Introduction

Various susceptibility genes and chromosomal abnormalities have been associated with autism spectrum disorder (ASD), but most discoveries either failed replication or accounted for a small effect. Although, it is widely accepted that multiple genes (and environmental factors) may contribute to the etiology of autism, the question of how many genetic susceptibility factors are involved and how they may be related cumulatively to the respective phenotypic subsets of ASD still remains unsolved. One of the main contributing factors to the low reproducibility rate in autism genetic studies is the lack of sufficient connectivity and integration among research findings. One way to address this gap is by employing integrated system biology-oriented research protocols. To demonstrate the importance of such approaches, a number of successfully applied methods are discussed here. The selected examples describe integrating different layers of genetic and epigenetic factors, gender disparity, and phenotypic subject stratification in autism studies as well as in bioinformatics pipelines thereby connecting the dots, which is expected to lead to the construction of autism-specific system biology networks (i.e., autism interactomes).

Whole Genome Studies

Numerous candidate gene and genome-wide association (GWAS) studies have been reported for autism with the goal of detecting common genetic risk variants, but with inconclusive findings. For instance, GWAS on two independent data sets, including family-based (n = 943 families) and case-control (n = 1,204 and 6,491 subjects, respectively) autism cohorts, did not detect genome-wide significant association initially. However, application of a meta-analysis approach recommended by de Bakker *et al.*[1] on both of

these independent data sets revealed genome-wide significance with the rs4307059 SNP on 5p14.1.[2] Evidence for association of autism with 5p14.1 SNPs was also reported by Ma *et al.*[3]

While replication of 5p14.1 association was intriguing, the remaining challenge was to uncover a potential role of this chromosomal region in the etiology of autism. The associated SNPs on 5p14.1 do not reside within known genes; therefore, assessment of the potential candidate genes (i.e., cadherin genes) did not provide conclusive results.[2] While conventional candidate gene screening was not fruitful, using a combination of bioinformatics analysis and expression studies Kerin *et al.*[4] discovered that a large noncoding RNA is transcribed from this autism associated region. Surprisingly, this product was encoded by a pseudogene (*MSNP1AS*), which was highly identical and antisense to the protein coding gene (*MSN*) located on the X chromosome. Individuals who carry the autism associated rs4307059 allele showed an elevation in the expression of this pseudogene. *MSNP1AS* can bind to the complementary transcript on the X chromosome resulting in down regulation of *MSN*. This innovative approach shed light on the interconnected nature of genetic factors and the pivotal role that understudied elements, such as noncoding RNAs and pseudogenes, may play in filling the gaps and unveiling the underlying pathogenic mechanisms of this complex condition.

Weiss *et al.*[5] conducted another large GWAS study using a combination of the transmission disequilibrium test (TDT) and case-control approaches, but found no support for association with the reported SNP at the 5p14.1 locus; instead, their study provided an evidence for association with an intergenic SNP near *SEMA5A* on chromosome 5p15. Initial gene expression studies provided evidence for reduced expression of *SEMA5A* in lymphoblastoid cell line (LCL), blood and brain samples in subjects with autism compared with controls.[5,6] However, a low minor allele frequency (MAF < 0.05) of the associated 5p15 SNP suggested that factors besides reduced expression of this single flanking gene might be involved in susceptibility to autism.

Later, the Autism Genome Project (AGP) consortium, representing more than 50 centers in North America and Europe, provided the largest family-based GWAS study to date, which included 1558 family trios.[7] No evidence was found for association with either 5p14.1 or 5p15 loci, but one SNP located at 20p12.1 (rs4141463 in a *MACROD2* intron) reached genome-wide association significance. However, a later attempt to replicate this finding in an independent case-control association study failed.[8] This lack of replication reflects the fundamental obstacle introduced by the extensive heterogeneity of ASD, which may even increase when subjects are recruited/compiled from diverse genetic backgrounds (e.g., ethnicities).

A follow-up *in silico* eQTL approach using a combination of expression and genotyping data sets revealed that rare variants in the *SEMA5A* regulatory network, including previous autism candidate genes and regions, such as *MACROD2*, may impact autism risk. This integrated approach provided evidence that the same pathway may contain both rare and common autism susceptibility genetic variants.[9] Therefore, collective assessments of different types of susceptibility factors enabled the detection of a potential connectivity between previously scattered findings.

Recent studies provided new insight into interpreting GWAS data and understanding genetic architecture of complex conditions, such as *cis* regulatory gene expression SNPs, by enrichment for expression quantitative trait loci (eQTL) and methylation QTL (mQTL) in relevant tissues.[10,11] Davis and colleagues[12] used existing GWAS data generated by the AGP,[7] genome-wide expression data sets in brain from Gamazon *et al.*,[11] and the SNP and CNV annotation databases.[13] This group adopted the approach previously used on bipolar disorder[11] to test for enrichment of identified eQTLs (in brain and LCLs) among top susceptibility SNPs from autism GWAS reports. This approach showed a global enrichment of brain eQTL (but not LCL eQTL) among top autism associated SNPs previously identified by GWAS.[7] Application of this annotation-based approach[12] also identified three genes as being strongly implicated in autism: *PANX1*, *PANX2*, and *SLC25A12*, two of which (*PANX2*, and *SLC25A12*) were previously reported to be differentially expressed in autism.[14] This type of research approach shows the direct and positive impact that integrated approaches provide for increasing the power of GWAS studies on neurodevelopmental disorders.

In addition to the necessity of providing functional perspective for susceptibility factors as discussed above, another obstacle in detecting genome-wide associated genes/alleles is the additive effect of other risk factors (i.e., genes, gender, and environmental factors), requiring assessment of interactions among multiple elements.[15] Because of the sex bias in autism, gender is one of the secondary factors that need to be assessed in GWAS studies. To improve detection power in autism GWAS studies, Lu and Cantor[15] conducted joint association tests[16] in a cohort of 990 multiplex autism families from the Autism Genetic Resource Exchange (AGRE) by incorporating gender as a risk factor in data analysis. The rationale behind this approach was to investigate the potential gender-specific differential penetrance of some ASD risk alleles. Stratification of the families into two groups based on the sex of the affected member [i.e., male-only (MO) and female-containing (FC)] revealed two associations exceeding genome-wide significance. The two susceptibility intronic alleles, rs6683048 (within *RyR2*) and rs1740138 (within *UPP2*), exhibited an over-transmission in both stratified groups (74% and 75% in the MO, 66% and 77% in the FC families, respectively). Although, the association was observed in both groups, it would not have been detected in a standard family based association test (i.e., without considering the sex factor) using the genome-wide level of significance.

More recent advancements in genomics technology provided a high throughput method for sequencing the entire protein coding regions of the genome using whole exome sequencing (WES). Detection of rare *de novo* loss of function (LoF) mutations has been the central point of most WES reports. However, considering the disease complexity and involvement of multiple genes, other types of *de novo* mutations as well as inherited events need to be assessed, in addition to LoF mutations, but they have been usually

excluded in conventional autism WES studies.[17] To address this critical gap, He *et al.*[17] developed a novel integrated statistical method, called TADA for "transmission and *de novo* association". This method yields gene-specific p-values and utilizes LoF *de novo* mutations in conjunction with inherited variations to identify autism risk genes. Such a method also allowed for incorporation of data from family-based and case-control studies, not possible in conventional analytical methods. Application of the TADA method to previously published WES data from about 1000 ASD trios[18–21] and 2000 subjects from a case-control study[22] demonstrated a dramatically higher detection power of this method compared with the commonly used methods (i.e., assessment of *de novo* LoF). Furthermore, some of the detected genes met genome-wide significance in this integrated model, while such a finding would be impossible to obtain without combining the *de novo*, transmitted and case-control data. This improvement in analyzing sequencing data illustrates that, unless used in conjunction with integrated methods, even the most advanced genomic methods (e.g., WES) would not be able to uncover the full complex picture of autism pathogenesis.

Gender Bias in Autism

Autism has a male preponderance of at least 4:1.[23] However, the biologic underpinnings of this gender discrepancy have not been fully elucidated. Among the most prominent proposed theories is the role of X-linked genes and epigenetic factors such as imprinting and X chromosome inactivation.[24] One of the first proposed theories to explain the increased ASD risk in males was a protective role of the X chromosome.[25] A premature speculation was that ASD might be an X-linked disorder, in which case females would have a protective mechanism provided by their second X chromosome.[26] Despite the emerging evidence in favor of a multi-factorial inheritance of ASD (ruling out an X-linked mode of transmission), prenatal sex hormone effects on autistic traits,[27–29] some compensatory effect mechanisms through sex chromosomal gene dosage or sex

hormone levels[30] may still exist. Additional area for consideration in relation to the X chromosome theory is the role of the X-linked microRNAs. The excessively high number (~10%) of all human microRNAs located on X versus autosomal chromosomes and their role in immune functions has provided support for another emerging hypothesis (i.e., an immunological advantage of females over males with respect to disease etiology).[31] Therefore, the degree to which X-linked microRNAs may account for the sex differences in autism should also be assessed.

Another proposed explanation for the gender bias in the prevalence of autism is a theory that males on average require fewer genetic risk factors necessary to reach an equivalent impairment threshold.[19,32] This hypothesis was tested from different perspectives; for instance, one study[33] demonstrated that in simplex autistic families, females were more severely affected (based on nonverbal IQ) than boys. This finding was not seen in multiplex families, which suggests the existence of underlying differences with respect to gender and impairment severity in simplex and multiplex families. The study conducted by Szatmari *et al.*[34] tested whether there are gender differences in severity of autism traits in multiplex families (divided into MO and FC) by focusing on repetitive behavior. This study demonstrated that in FC families autistic male relatives are more severely affected than in MO families, a finding also supported by the largest (n = 192 pairs) population based autism twin study.[35] These observations were replicated in a recent study,[36] further supporting the above discussed hypothesis of a female protective effect against autism.

Even though contributing factors to the observed gender bias still need to be elucidated, the emerging evidence has shown that implementing gender-base stratifications in autism genetic studies will increase detection power. This increase in power has been shown in SNP and candidate gene association studies, as well as in the detection of chromosomal abnormalities. For example, copy number variations at the chromosome 16p13 region have been frequently associated with neurodevelopmental disorders, including autism. Assessment of 46 cases with chromosome 16p13 CNVs

(both deletions and duplications) showed a sex-biased effect, with a significant enrichment of CNVs only in males, but not in females.[37] This finding demonstrates that taking into account the gender bias is crucial in investigating the potential pathological role of the autism associated CNVs.

By applying a network-based analysis on genome-wide *de novo* CNV data[32] generated from a large cohort of simplex families Gilman *et al.*[38] tested a hypothesis that underlying disease-causing genetic disturbances in females were much stronger than in males. This analysis provided support for two contributing mechanisms: (1) significantly larger CNVs were observed in females than males (harboring about 10 and 3 genes per CNV, respectively) and (2) genes affected by CNVs in females had higher impact on the overall network cluster score than those in males.

In addition to CNVs, the impact of gender bias has also been shown in other autism genetic studies, including linkage and GWAS. In one of the earlier whole genome scans addressing sex bias, a linkage study was conducted on the basis of the sex of affected individuals (i.e., MO and FC families), which identified a male-specific linkage peak.[39] The importance of the applied gender stratification in this linkage study is further highlighted by the fact that the original non-stratified genome scan[40] failed to differentiate the autistic cohort from the general population. Applying sex as an additional risk factor in GWAS of multiplex autism families enhanced the study power and resulted in the detection of sex-specific genetic variants of genome-wide significance.[15]

Not only did inclusion of gender effect increase the detection power in finding autism susceptibility factors, it also improved autism risk assessment models. Carayol *et al.*[41] improved their previously proposed risk assessment genetic scoring system[42] by developing sex-specific score models to identify siblings with significantly higher risk of autism. This approach identified specific sets of SNPs with a high reproducibility index in males and females. Comparing the previous model[42] with this sex-specific scoring system showed a higher detection power in the latter.

Phenotypic Stratification

Autism is often accompanied by other symptom or comorbidity; such co-occurring symptoms may serve as biomarkers to help identify subsets with a greater chance of having an underlying mechanism that plays a role in both ASD and the concurrent non-ASD phenotypes. These co-occurring symptoms have been used in subject stratification for genetic linkage, association, and mutation screening studies. The majority of the findings for these stratified ASD subsets have been successfully replicated, further demonstrating the value of implementing stratification methods for better understanding of the genetic etiology of this heterogeneous condition. Two highly replicated discoveries of causative genes in autism (*PTEN* and *MET*) were made by focusing on co-occurring symptoms (i.e., macrocephaly and gastrointestinal dysfunction, respectively) in subsets of autistic subjects.[43–46]

In addition to co-occuring symptoms, behavioral profiles may be useful in narrowing down the existing phenotypic heterogeneity in autism. Since individuals with ASD manifest deficiencies across a broad range of behaviors, addressing phenotypic heterogeneity requires more inclusive methods than subject stratification based on severity along a single domain, such as language impairment or nonverbal communication. In an attempt to address this gap, Hu and Steinberg[47] applied multiple clustering methods on more than 100 ADI-R scores across a broad range of symptoms, which identified four subgroups of autistic individuals representing distinct phenotypes. Heterogeneity in ASD is also reflected at the family level and, thus, in multiplex families autistic symptoms may vary among affected siblings. Therefore, to address inter- and intra-family heterogeneity we developed a multi-step subject stratification approach[48] by incorporating the proposed ADI-R classification[47] on multiplex families.

To test the detection power of this model in linkage scans, we re-analyzed existing SNP genotyping data for 392 multiplex AGRE families from published studies.[5,49] The affected subject's ADI-R

sub-phenotype[47] was used to create group-specific SNP data sets using a multi-step stratification process. The applied deep stratification method provided a pipeline to further filter the original heterogeneous ASD pedigree data file into more homogeneous data sets by first using ADI-R cluster analysis, followed by removal of sub-phenotypically discordant siblings, and finally by separation of MO and FC pedigrees.[48]

No locus reached significance for the combined non-stratified cohort (n = 392 families). However, study-wide significant linkage scores (i.e., LOD score greater than 4) were reached for chromosomes 22q11 and 13q21 for two ASD subsets (n = 232 and 63 families, respectively) representing the most severely language-impaired individuals. Notably, 13q21 has been previously linked to autism with language impairment, and 22q11 has been separately associated with either autism or language disorders. Linkage analysis on chromosome 5p15 for a combination of two stratified FC subgroups (n = 23 families) demonstrated suggestive linkage (LOD = 3.5), which replicates previous linkage results for FC pedigrees.[49] A trend was also found for the association of previously reported autism associated 5p14-p15 SNPs[2,5] in the same combined FC cohort.

Interestingly, conducting a similar analysis on all FC families (n = 166), without applying the ADI-R subtyping, did not detect associations to this chromosomal region. Our findings in a given ADI-R stratified FC cohort not only confirmed previous linkage[49] and association[2] reports, but also further refined the proposed link between the 5p region and autism. Furthermore, examination of the associated noncoding RNA (*MSNP1AS*) in the identified homogeneous FC cohort in our study (that showed positive associations to the susceptibility SNPs at 5p) may enhance the revelation of other autism risk factors in the intriguing model proposed by Kerin *et al.*[4]

This integrated approach illustrated the added likelihood of detecting significant linkage when the heterogeneity of the ASD population is reduced by proper sample stratification. It also provided first lines of evidence at the gene scan level for both inter- and intra-family heterogeneity, reflecting both shared and distinct genetic makeup in the autism population. Inter- and intra-family heterogeneity is another important factor that has been overlooked in whole-genome screening of multiplex families.

Network Analysis

It has been more than a decade since researchers became interested in developing autism network models to predict gene/environment or gene/gene interactions by evaluating experimental findings using computational tools and data-mining. In one of the earliest papers on this topic[50] multiple bioinformatics tools, including the Pathwayassist software (Ariadne Genomics, Rockville, MD) and the ResNet database, were applied to positional autism candidate genes identified by an earlier genome-wide linkage study.[40] The objectives of the data-mining study were to prioritize candidate genes, predict pathways of interacting genes, and identify biologically meaningful autism candidates. Since then, a wider spectrum of analytical tools have been developed and applied to extensive autism genetic data, and some of them will be highlighted here.

Lee *et al.*[51] performed an integrative autism gene network analysis based on seven phenotypic features commonly reported in autism, in conjunction with multiple sets of autism genetic data (i.e., GWAS, CNV, and gene expression). Genes related to these seven restricted phenotypes were retrieved from the curated databases (Ingenuity, GeneGo, and HuGe navigator). A novel CNV-centric node network analysis method, as well as GWAS and expression data were then used to identify potential regulatory networks in autism.

Later, literature-based curated data for autism-specific rare and syndromic candidate genes were used to build a composite reference profile based on functional and expression analyses using an integrated bioinformatics approach.[52] A brain region-specific predictive gene map for autism was built by creating a functional profile of the reference set of ASD candidate genes using GO enrichment analysis, along with an expression profile utilizing DAVID analysis. The functional profile identified enrichment of

three biological processes critical for synaptic transmission in ASD genes. Such a predictive gene map may facilitate prioritizing autism-related genomic data.

Poelmans *et al.*[29] constructed protein networks by conducting bioinformatics analysis and integrating SNP data from six GWAS studies.[2,3,5,7,53,54] This finding was further corroborated by obtaining evidence from exome sequencing, CNV, expression studies (gene and microRNA), as well as genetic animal studies, which resulted in the identification of three ASD-related signaling networks. Furthermore, this bioinformatics and literature analyses data integration study showed overlap between rare and common ASD implicated genetic variants.

More recently, to examine the multi-factorial nature of autism, Zeidan-Chulia and colleagues[55] developed a computational method using a combination of thorough literature search and a system biology approach. One of their objectives was to determine the main biological processes and the dominant contributing factors (genetic and/or environmental) integral to the disorder. Such *in silico* models may offer a way to identify potentially relevant genetic and environmental factors as well as their interactions in autism.

Biological relevance of disease associated genes needs to be confirmed by functional experiments. One of the challenges in pursuing productive functional studies for autism genes is to know the relevant spatial and temporal gene expression pattern in brain. To address this gap, Willsey *et al.*[56] constructed co-expressed networks based on a list of high confidence ASD genes identified by exome and genome-wide sequencing studies and using a publicly available spatio-temporal transcriptome of the human brain data set[57] as the substrate. Such an integrated analytic approach provided a potential basis for identification of pathophysiological mechanisms of ASD-related mutations by guiding when, where, and in what cell type should a candidate gene be studied.[56]

Interactome Networks

The importance of employing integrated approaches in understanding the pathophysiology of human diseases has been further highlighted in the genetic field by introducing a new concept of interactome networks. This concept refers to the interactions between biological systems and complex cellular networks, and has been proposed as a next step toward predicting genotype-phenotype related outcomes.[58] There are different types of interactome networks in cellular systems, including metabolic, protein-protein interaction, gene regulatory systems, and networks of functional links (transcriptional profiling, phenotypic and genetic interaction networks). To avoid "simplifying" compound systems and losing the functional influence of each component, assessment of a combination of networks is essential. Because of the overwhelming intricacy of biological network perturbations within cells, the interactome network mapping has not yet been fully completed; however, mapping has been compiled for several model organisms and more advancements are underway for human as well.[58]

As anticipated, the concept of interactomes has begun to receive attention in the autism research community as shown by the development of an autism-specific protein interactome.[59] Such efforts can be further expanded by incorporating other networks, such as microRNA inter-regulatory human interactomes.[60] Integrated approaches in autism genetic studies such as those discussed in this chapter will provide additional knowledge-base for advancing autism interactomes re-construction.

Future Directions

Despite a growing recognition of the multi-factorial architecture of autism and the need to consider gender disparity and phenotypic heterogeneity, these essential elements are still often overlooked in study designs and research protocols. A substantial portion of the research funding has already been devoted to generating the rapidly growing amount of different type of genomic data from autistic subjects. Autism initiatives (i.e., the Autism Genetic Resource Exchange (AGRE), Simons Simplex Collection (SSC), and the National Database for Autism Research (NDAR)[61]) and consortiums (i.e., Autism Genome Project (AGP) and Autism Sequencing)

have provided invaluable resources to the research community by collecting phenotypic data and supporting the ongoing generation of genomic data (e.g., genotyping, CNV, sequencing, and expression) from subjects with autism.

A long-needed shift from generating new data to re-assessment of already existing phenotypic and genomic data along with developing novel integrated research models would accelerate the uncovering of previously undetected autism risk factors. If incorporated in genomic data re-assessments, established concepts such as sex bias, comorbidity, and phenotypic heterogeneity have a great potential to accelerate replication of findings, subtype detection, and biomarker identification, as well as to facilitate the development of evidence-based and more effective treatment strategies for autism.

Executive Summary

- Multiple genes (and environmental factors) may contribute to the etiology of non-oligogenic forms of autism, which necessitates cumulative assessment of potential genetic and epigenetic contributing factors.
- Integrated approaches (i.e., utilizing two or more phenotypic and genetic factors as well as bioinformatics and data mining pipelines) would be beneficial in detecting biologically relevant and replicated findings that move the autism genetic field forward.
- The examples discussed in this chapter illustrate how connectivity among multiple lines of scattered autism research findings can be established by employing integrated system biology-oriented research protocols.
- Abnormalities in gene regulatory processes such as noncoding RNAs, alternative splicing, and DNA methylation may play a critical role in pathogenesis of complex conditions. The existing and ongoing large-scale autism genetic sequencing and expression data need to be also analyzed with respect to assessing the gene regulatory factors.

- Considering the extensive crosstalk between genetic and epigenetic mechanisms assessment of cumulative and interactive effects of multiple factors in autism requires designing comprehensive research plans and data analysis strategies to unveil pathogenesis of multi-factorial forms of ASD.

References

1. de Bakker PI, Ferreira MA, Jia X, Neale BM, Raychaudhuri S, Voight BF: Practical aspects of imputation-driven meta-analysis of genome-wide association studies. *Hum Mol Genet* 2008, **17**(R2):R122–128.
2. Wang K, Zhang H, Ma D, Bucan M, Glessner JT, Abrahams BS, Salyakina D, Imielinski M, Bradfield JP, Sleiman PM *et al.*: Common genetic variants on 5p14.1 associate with autism spectrum disorders. *Nature* 2009, **459**(7246):528–533.
3. Ma D, Salyakina D, Jaworski JM, Konidari I, Whitehead PL, Andersen AN, Hoffman JD, Slifer SH, Hedges DJ, Cukier HN *et al.*: A genome-wide association study of autism reveals a common novel risk locus at 5p14.1. *Ann Hum Genet* 2009, **73**(Pt 3):263–273.
4. Kerin T, Ramanathan A, Rivas K, Grepo N, Coetzee GA, Campbell DB: A noncoding RNA antisense to moesin at 5p14.1 in autism. *Sci Transl Med* 2012, **4**(128):128ra140.
5. Weiss LA, Arking DE, Daly MJ, Chakravarti A: A genome-wide linkage and association scan reveals novel loci for autism. *Nature* 2009, **461**(7265): 802–808.
6. Melin M, Carlsson B, Anckarsater H, Rastam M, Betancur C, Isaksson A, Gillberg C, Dahl N: Constitutional downregulation of SEMA5A expression in autism. *Neuropsychobiology* 2006, **54**(1):64–69.
7. Anney R, Klei L, Pinto D, Regan R, Conroy J, Magalhaes TR, Correia C, Abrahams BS, Sykes N, Pagnamenta AT *et al.*: A genome-wide scan for common alleles affecting risk for autism. *Hum Mol Genet* 2010, **19**(20):4072–4082.
8. Curran S, Bolton P, Rozsnyai K, Chiocchetti A, Klauck SM, Duketis E, Poustka F, Schlitt S, Freitag CM, Lee I *et al.*: No association between a common single nucleotide polymorphism, rs4141463, in the MACROD2 gene and autism spectrum disorder. *Am J Med Genet B Neuropsychiatr Genet* 2011, **156B**(6):633–639.
9. Cheng Y, Quinn JF, Weiss LA: An eQTL mapping approach reveals that rare variants in the SEMA5A regulatory network impact autism risk. *Hum Mol Genet* 2013, **22**(14):2960–2972.
10. Nicolae DL, Gamazon E, Zhang W, Duan S, Dolan ME, Cox NJ: Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS. *PLoS Genet* 2010, **6**(4):e1000888.

. Gamazon ER, Badner JA, Cheng L, Zhang C, Zhang D, Cox NJ, Gershon ES, Kelsoe JR, Greenwood TA, Nievergelt CM *et al.*: Enrichment of cis-regulatory gene expression SNPs and methylation quantitative trait loci among bipolar disorder susceptibility variants. *Mol Psychiatry* 2013, **18**(3):340–346.

2. Davis LK, Gamazon ER, Kistner-Griffin E, Badner JA, Liu C, Cook EH, Sutcliffe JS, Cox NJ: Loci nominally associated with autism from genome-wide analysis show enrichment of brain expression quantitative trait loci but not lymphoblastoid cell line expression quantitative trait loci. *Mol Autism* 2012, **3**(1):3.

3. Gamazon ER, Zhang W, Konkashbaev A, Duan S, Kistner EO, Nicolae DL, Dolan ME, Cox NJ: SCAN: SNP and copy number annotation. *Bioinformatics* 2010, **26**(2):259–262.

4. Voineagu I, Wang X, Johnston P, Lowe JK, Tian Y, Horvath S, Mill J, Cantor RM, Blencowe BJ, Geschwind DH: Transcriptomic analysis of autistic brain reveals convergent molecular pathology. *Nature* 2011, **474**(7351):380–384.

5. Lu AT, Cantor RM: Allowing for sex differences increases power in a GWAS of multiplex Autism families. *Mol Psychiatry* 2012, **17**(2):215–222.

6. Cordell HJ, Barratt BJ, Clayton DG: Case/pseudocontrol analysis in genetic association studies: a unified framework for detection of genotype and haplotype associations, gene-gene and gene-environment interactions, and parent-of-origin effects. *Genet Epidemiol* 2004, **26**(3):167–185.

7. He X, Sanders SJ, Liu L, De Rubeis S, Lim ET, Sutcliffe JS, Schellenberg GD, Gibbs RA, Daly MJ, Buxbaum JD *et al.*: Integrated model of de novo and inherited genetic variants yields greater power to identify risk genes. *PLoS Genet* 2013, **9**(8):e1003671.

8. Sanders SJ, Murtha MT, Gupta AR, Murdoch JD, Raubeson MJ, Willsey AJ, Ercan-Sencicek AG, DiLullo NM, Parikshak NN, Stein JL *et al.*: De novo mutations revealed by whole-exome sequencing are strongly associated with autism. *Nature* 2012, **485**(7397):237–241.

9. Neale BM, Kou Y, Liu L, Ma'ayan A, Samocha KE, Sabo A, Lin CF, Stevens C, Wang LS, Makarov V *et al.*: Patterns and rates of exonic de novo mutations in autism spectrum disorders. *Nature* 2012, **485**(7397):242–245.

0. O'Roak BJ, Vives L, Girirajan S, Karakoc E, Krumm N, Coe BP, Levy R, Ko A, Lee C, Smith JD *et al.*: Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. *Nature* 2012, **485**(7397):246–250.

1. Iossifov I, Ronemus M, Levy D, Wang Z, Hakker I, Rosenbaum J, Yamrom B, Lee YH, Narzisi G, Leotta A *et al.*: De novo gene disruptions in children on the autistic spectrum. *Neuron* 2012, **74**(2):285–299.

2. Liu L, Sabo A, Neale BM, Nagaswamy U, Stevens C, Lim E, Bodea CA, Muzny D, Reid JG, Banks E *et al.*: Analysis of rare, exonic variation amongst subjects with autism spectrum disorders and population controls. *PLoS Genet* 2013, **9**(4):e1003443.

23. Volkmar FR, Szatmari P, Sparrow SS: Sex differences in pervasive developmental disorders. *J Autism Dev Disord* 1993, **23**(4):579–591.

24. Marco EJ, Skuse DH: Autism-lessons from the X chromosome. *Soc Cogn Affect Neurosci* 2006, **1**(3):183–193.

25. Skuse DH: Imprinting, the X-chromosome, and the male brain: explaining sex differences in the liability to autism. *Pediatr Res* 2000, **47**(1):9–16.

26. Baron-Cohen S, Lombardo MV, Auyeung B, Ashwin E, Chakrabarti B, Knickmeyer R: Why are autism spectrum conditions more prevalent in males? *PLoS Biol* 2011, **9**(6):e1001081.

27. Auyeung B, Baron-Cohen S, Ashwin E, Knickmeyer R, Taylor K, Hackett G: Fetal testosterone and autistic traits. *Br J Psychol* 2009, **100**(Pt 1):1–22.

28. Auyeung B, Taylor K, Hackett G, Baron-Cohen S: Foetal testosterone and autistic traits in 18 to 24-month-old children. *Mol Autism* 2010, **1**(1):11.

29. Poelmans G, Franke B, Pauls DL, Glennon JC, Buitelaar JK: AKAPs integrate genetic findings for autism spectrum disorders. *Transl Psychiatry* 2013, **3**:e270.

30. Sarachana T, Xu M, Wu RC, Hu VW: Sex hormones in autism: androgens and estrogens differentially and reciprocally regulate RORA, a novel candidate gene for autism. *PLoS One* 2011, **6**(2):e17116.

31. Pinheiro I, Dejager L, Libert C: X-chromosome-located microRNAs in immunity: might they explain male/female differences? The X chromosome-genomic context may affect X-located miRNAs and downstream signaling, thereby contributing to the enhanced immune response of females. *Bioessays* 2011, **33**(11):791–802.

32. Levy D, Ronemus M, Yamrom B, Lee YH, Leotta A, Kendall J, Marks S, Lakshmi B, Pai D, Ye K *et al.*: Rare de novo and transmitted copy-number variation in autistic spectrum disorders. *Neuron* 2011, **70**(5):886–897.

33. Banach R, Thompson A, Szatmari P, Goldberg J, Tuff L, Zwaigenbaum L, Mahoney W: Brief Report: relationship between non-verbal IQ and gender in autism. *J Autism Dev Disord* 2009, **39**(1):188–193.

34. Szatmari P, Liu XQ, Goldberg J, Zwaigenbaum L, Paterson AD, Woodbury-Smith M, Georgiades S, Duku E, Thompson A: Sex differences in repetitive stereotyped behaviors in autism: implications for genetic liability. *Am J Med Genet B Neuropsychiatr Genet* 2012, **159B**(1):5–12.

35. Hallmayer J, Cleveland S, Torres A, Phillips J, Cohen B, Torigoe T, Miller J, Fedele A, Collins J, Smith K *et al.*: Genetic heritability and shared environmental factors among twin pairs with autism. *Arch Gen Psychiatry* 2011, **68**(11):1095–1102.

36. Robinson EB, Lichtenstein P, Anckarsater H, Happe F, Ronald A: Examining and interpreting the female protective effect against autistic behavior. *Proc Natl Acad Sci U S A* 2013, **110**(13):5258–5262.

37. Tropeano M, Ahn JW, Dobson RJ, Breen G, Rucker J, Dixit A, Pal DK, McGuffin P, Farmer A, White PS *et al.*: Male-biased autosomal effect of

16p13.11 copy number variation in neurodevelopmental disorders. *PLoS One* 2013, **8**(4):e61365.

38. Gilman SR, Iossifov I, Levy D, Ronemus M, Wigler M, Vitkup D: Rare de novo variants associated with autism implicate a large functional network of genes involved in formation and function of synapses. *Neuron* 2011, **70**(5):898–907.
39. Stone JL, Merriman B, Cantor RM, Yonan AL, Gilliam TC, Geschwind DH, Nelson SF: Evidence for sex-specific risk alleles in autism spectrum disorder. *Am J Hum Genet* 2004, **75**(6):1117–1123.
40. Yonan AL, Alarcon M, Cheng R, Magnusson PK, Spence SJ, Palmer AA, Grunn A, Juo SH, Terwilliger JD, Liu J *et al.*: A genomewide screen of 345 families for autism-susceptibility loci. *Am J Hum Genet* 2003, **73**(4):886–897.
41. Carayol J, Schellenberg GD, Dombroski B, Genin E, Rousseau F, Dawson G: Autism risk assessment in siblings of affected children using sex-specific genetic scores. *Mol Autism* 2011, **2**(1):17.
42. Carayol J, Schellenberg GD, Tores F, Hager J, Ziegler A, Dawson G: Assessing the impact of a combined analysis of four common low-risk genetic variants on autism risk. *Mol Autism* 2010, **1**(1):4.
43. Butler MG, Dasouki MJ, Zhou XP, Talebizadeh Z, Brown M, Takahashi TN, Miles JH, Wang CH, Stratton R, Pilarski R *et al.*: Subset of individuals with autism spectrum disorders and extreme macrocephaly associated with germline PTEN tumour suppressor gene mutations. *J Med Genet* 2005, **42**(4):318–321.
44. McBride KL, Varga EA, Pastore MT, Prior TW, Manickam K, Atkin JF, Herman GE: Confirmation study of PTEN mutations among individuals with autism or developmental delays/mental retardation and macrocephaly. *Autism Res* 2010, **3**(3):137–141.
45. Herman GE, Butter E, Enrile B, Pastore M, Prior TW, Sommer A: Increasing knowledge of PTEN germline mutations: two additional patients with autism and macrocephaly. *Am J Med Genet A* 2007, **143**(6):589–593.
46. Campbell DB, Buie TM, Winter H, Bauman M, Sutcliffe JS, Perrin JM, Levitt P: Distinct genetic risk based on association of MET in families with co-occurring autism and gastrointestinal conditions. *Pediatrics* 2009, **123**(3):1018–1024.
47. Hu VW, Steinberg ME: Novel clustering of items from the Autism Diagnostic Interview-Revised to define phenotypes within autism spectrum disorders. *Autism Res* 2009, **2**(2):67–77.
48. Talebizadeh Z, Arking DE, Hu VW: A novel stratification method in linkage studies to address inter- and intra-family heterogeneity in autism. *PLoS One* 2013, **8**(6):e67569.
49. Szatmari P, Paterson AD, Zwaigenbaum L, Roberts W, Brian J, Liu XQ, Vincent JB, Skaug JL, Thompson AP, Senman L *et al.*: Mapping autism risk loci using genetic linkage and chromosomal rearrangements. *Nat Genet* 2007, **39**(3):319–328.

50. Yonan AL, Palmer AA, Smith KC, Feldman I, Lee HK, Yonan JM, Fischer SG, Pavlidis P, Gilliam TC: Bioinformatic analysis of autism positional candidate genes using biological databases and computational gene network prediction. *Genes Brain Behav* 2003, **2**(5):303–320.
51. Lee TL, Raygada MJ, Rennert OM: Integrative gene network analysis provides novel regulatory relationships, genetic contributions and susceptible targets in autism spectrum disorders. *Gene* 2012, **496**(2):88–96.
52. Kumar A, Swanwick CC, Johnson N, Menashe I, Basu SN, Bales ME, Banerjee-Basu S: A brain region-specific predictive gene map for autism derived by profiling a reference gene set. *PLoS One* 2011, **6**(12):e28431.
53. Salyakina D, Ma DQ, Jaworski JM, Konidari I, Whitehead PL, Henson R, Martinez D, Robinson JL, Sacharow S, Wright HH *et al.*: Variants in several genomic regions associated with asperger disorder. *Autism Res* 2010, **3**(6):303–310.
54. Hussman JP, Chung RH, Griswold AJ, Jaworski JM, Salyakina D, Ma D, Konidari I, Whitehead PL, Vance JM, Martin ER *et al.*: A noise-reduction GWAS analysis implicates altered regulation of neurite outgrowth and guidance in autism. *Mol Autism* 2011, **2**(1):1.
55. Zeidan-Chulia F, Rybarczyk-Filho JL, Salmina AB, de Oliveira BH, Noda M, Moreira JC: Exploring the multifactorial nature of autism through computational systems biology: calcium and the Rho GTPase RAC1 under the spotlight. *Neuromolecular Med* 2013, **15**(2):364–383.
56. Willsey AJ, Sanders SJ, Li M, Dong S, Tebbenkamp AT, Muhle RA, Reilly SK, Lin L, Fertuzinhos S, Miller JA *et al.*: Coexpression networks implicate human midfetal deep cortical projection neurons in the pathogenesis of autism. *Cell* 2013, **155**(5):997–1007.
57. Kang HJ, Kawasawa YI, Cheng F, Zhu Y, Xu X, Li M, Sousa AM, Pletikos M, Meyer KA, Sedmak G *et al.*: Spatio-temporal transcriptome of the human brain. *Nature* 2011, **478**(7370):483–489.
58. Vidal M, Cusick ME, Barabasi AL: Interactome networks and human disease. *Cell* 2011, **144**(6):986–998.
59. Sakai Y, Shaw CA, Dawson BC, Dugas DV, Al-Mohtaseb Z, Hill DE, Zoghbi HY: Protein interactome reveals converging molecular pathways among autism disorders. *Sci Transl Med* 2011, **3**(86):86ra49.
60. Sengupta D, Bandyopadhyay S: Participation of microRNAs in human interactome: extraction of microRNA-microRNA regulations. *Mol Biosyst* 2011, **7**(6):1966–1973.
61. Hall D, Huerta MF, McAuliffe MJ, Farber GK: Sharing heterogeneous data: the national database for autism research. *Neuroinformatics* 2012, **10**(4):331–339.