The genetic background of ASD in a Hungarian ASD cohort

PhD thesis

Noémi Ágnes Varga MD

János Szentágothai Doctoral School of Neurosciences
Semmelweis University

Supervisor: Mária Judit Molnár, MD, DSc
Official reviewers: Krisztina Vellainé Takács, PhD
Bálint Szuromi, MD, PhD
Head of the Final Examination Comittee: Viktor Farkas, MD, PhD
Members of the Final Examination Comittee:
Ildikó Sipos, MD, PhD
Zoltán Liptai, MD, PhD

Budapest
2020

INTRODUCTION

The genetic architecture of autism spectrum disorder

Autism spectrum disorder (ASD) is a neurodevelopmental disorder, characterized by the core symptoms of impaired social communication, restricted interests and stereotyped, repetitive behavior. Beside its phenotypical heterogeneity the ethiology is also very complex and heterogenous. ASD has an estimated heritability of 64–91%, suggesting a strong genetic effect. Former genetic studies suggested that both common and rare genetic variants play a role in the etiology, and mutation types range from single nucleotide variants to large chromosomal aberrations, as well as variations in regulatory DNA elements. The number of genes implicated in ASD pathogenesis is >1,000 according to the SFARI database, which converge on different cellular pathways. Genes implicated in synaptic strength, cell growth and differentiation, as well as genes involved in chromatin remodeling where discovered as candidate genes in pathogenesis of ASD. A shared genetic environment is possible with other psychiatric disorders

as well, especially with schizophrenia. The clinical expression of ASD is nevertheless varied, despite the common umbrella term. This high phenotypic variability, however, is mirrored poorly in the everyday clinical diagnostics of ASD, which doesn't serve the analysis of the genetic-phenotypic background. Earlier attempts to identify ASD subtypes or endophenotypes based on clinical features have met with limited success, and there were attempts to define subtypes based on genetics (endophenotypes), which led to identify several monogenic ASD cases, as *CHD8* or *AUTS2* subtype. According to the classical definition, the syndromic ASD is a "disorder with a clinically defined pattern of somatic abnormalities and a neurobehavioral phenotype that may include ASD," however, this is only present in cca. 10% of cases. Instead, if we move to a molecularly defined approach, in 25% of the cases there is a detectable strong genetic change, which may be even higher if multiple minor physical anomalies (MPA) are present.

Mitochondrial dysfunction has been described in ASD; however, primary mitochondrial disease has been

genetically proven in a small subset of patients. Despite evidence of altered mitochondrial function in some individuals with ASD, it is not known whether mitochondrial dysfunction is a cause or an effect of ASD.

Mitochondrial dysfunction can be caused by either primary mitochondrial disease (MD) or by secondary mitochondrial damage. In case of primary MD pathogenic mutations are found in mitochondrial DNA (mtDNA) or nuclear genes involved in directly in mitochondrial function. Secondary mitochondrial dysfunction could be caused by nongenetic-environmental causes or from mutations in genes encoding neither function nor production of the oxidative phosphorylation (OXPHOS) proteins. Furthermore, mitochondrial dysfunction has been associated with some forms of syndromic ASD but also in cases of non-syndromic ASD.

OBJECTIVES

Our objectives were the followings:

1. To detect and analyze the rare, monogenic causes of ASD in Hungarian cohort.
2. To investigate the significance of rare variants on a cohort level.
3. To confirm whether rare variant burden associates with autism severity.
4. To study the presence of autism phenotypic clusters, and if these clusters are associated with specific rare variants.
5. To confirm the presence of a mitochondrial ASD subgroup, and to elucidate the etiology of the mtDNA deletion in ASD.

METHODS

Patients involved in genetic studies

174 ASD patients were recruited from the Vadaskert Child and Adolescent Psychiatry Hospital and Outpatient Clinic. Detailed clinical examinations consisting of a general medical examination and neurological assessment were performed. A diagnosis of ASD was made by a qualified psychologist using the ADI-R (Autism Diagnostic Interview-Revised) and ADOS (Autism Diagnostic Observation Schedule). Patients were screened for MPA, which were selected based on the Méhes Scale. Family history and detailed environmental/societal data were collected from the parent (or parents) of each patient. Any disorders present in the parents, as well as environmental factors, were registered. Regarding the investigation of mitochondrial dysfunction in ASD 60 ASD patients and 60 healthy individuals were screened for common mtDNA mutations. Cohorts of healthy controls, ASD patients without mtDNA alterations, and patients with primary mitochondrial disorders without ASD served as

comparison groups for the interpretation of our next generation sequencing (NGS) results.

Methods of genetic analysis

DNA was isolated from peripheral blood samples with QIAamp DNA blood kit. We performed Fragile X-screening at every patient using the Amplidex FMR1 PCR kit. All patients with Fragile-X-syndrome were excluded from all subsequent genetic analysis and phenotypic cluster analysis. The 101 ASD-associated genes were investigated with NGS, which was performed on a MiSeq platform using the TruSight Autism Rapid Capture Kit and the SureSelect QXT Kit. In the autism panel, 24 samples were multiplexed in a single run using the MiSeq reagent kit v2 and 300 cycles. The mean read depth was 135× in the gene panel, 20× coverage was achieved in a minimum of 90% of target regions. Pathogenic (P) and likely pathogenic (LP) mutations from NGS data were validated by Sanger sequencing. Parents were Sanger sequenced for specific variants in selected cases, where syndromic ASD was suspected based on the sequencing results of the index case, and the parent was

available for genetic analysis. To identify mtDNA deletions long range PCR was performed. NGS was performed on patients with major mtDNA deletions (mtdel-ASD) using two gene panels: the ASD panel and the intergenomical communication (IG) panel to investigate nuclear genes that are responsible for mtDNA maintenance.

Methods of statistical analysis

Chi square test with Yates correction/Fisher exact test were used to determine signifcant diferences between patient and control groups. Raw sequences were filtered with Picard tools and quality filtered reads were aligned to the hg19 reference genome with BWA-mem. Variant calling was performed using GATK HaplotypeCaller and VCF files were annotated with SnpEff. We analyzed only those variants that were found in the canonical transcripts of the gene. Variant quality was assessed by GATK, and only variants, which were flagged as PASS (RD >10, Mapping quality >40, quality by depth >2) were analyzed. To filter potentially causal single-gene Mendelian variations on a case-by-case level, we used the

VariantAnalyzer in-house software developed at the Budapest University of Technology and Economics. This software application annotates SNPs and short INDELs with several types of annotations, including the MAF from 1000 Genomes Project, ExAC, conservation scores based on GERP, predicted function of non-synonymous SNPs using dbNSFP, and disease associations with HGMD and ClinVar. First, we filtered for variants known to be disease-causing. Second, we filtered for rare variants based on the minor allele frequency (MAF) and frequency of the mutation in our 200 Whole Exome Sequencing (WES) repository. We labeled a variant as rare if it was present in one homozygous or two heterozygous samples within our cohort (equal to a MAF-cutoff of ~5‰), and the MAF in Europeans from the 1000 Genomes and ExAC databases, as well as in our in-house exome database was less than 5%. Mutations were prioritized based on their predicted effects. Exonic frameshifts, stop mutations and canonical splice site variants were considered damaging or loss of function (LoF), whereas the effects of missense mutations were predicted using multiple prediction tools (Polyphen2,

SIFT, etc.). The variants were assessed as recommended by the ACMG guideline 2015.

For the analysis of rare variants in a common disease-rare variant framework on a cohort level, first we tested, whether the total number of detected rare missense and LoF variants in a given gene is greater than expected. P-value was calculated with the associated software: SORVA. Secondly for the calculation of rare variant burden, genes were normalized according to genic intolerance to mutation. Specifically, we used the inverse RVIS percentile [1–(RVIS percentile÷100)], to give a weight to every gene than summed the number of variants in a given patient [Σ(variants × weightedgene-score)]. Linear regression was used then to test for correlation between rare variant burden and autism severity, and rare variant burden vs. MPA burden. Autism severity was assessed by the total ADOS score, in patients, at whom ADOS was available (N = 47), and also by calibrated severity score (CSS). For comparison of rare variant burden in males versus females, and the number of minor malformations in syndromic versus non-syndromic cases

two-tailed T-test was used. For the analysis of rare variant association with potential autism subphenotypes first, we assessed, whether such subphenotypes can be created based solely on the clinical data. We have used our clinical questionnaire and ADI-R results for cluster analysis. For the phenotypic cluster analysis we utilized kernel PCA and spectral clustering. In accordance with earlier studies, we set the number of clusters in spectral clustering to four. We also investigated the variables characterizing each cluster via computing the relative frequency of the presence of each feature. We performed the kernel PCA with 3 dimensions to visualize the transformed samples and cluster assignments. To define potential endophenotypes (subgroup with specific genetic background) the characterized four clusters where used and correlated with the detected rare variants aggregated on candidate genes. To assess the correlation between the subphenotypes and genetics, we investigated whether detected rare variants of a candidate gene occur more frequently in either of the resulting clusters using ANOVA and pairwise T-tests, in conjunction with Bonferroni correction for multiple hypothesis testing.

RESULTS

The detected syndromic ASD cases

We diagnosed 13 syndromic autism cases based on our genetic findings. Four patients were diagnosed with FXS, however, one of them is a girl. In this case, the family history (2 healthy brothers), did not suggest an X-linked inheritance. Overall eight mutations were considered as pathogenic according to ACMG classification, with close phenotype match (four patients with FXS, one with Dravet syndrome, one with CHARGE syndrome, one with Duchenne muscular dystrophy, one with atypical Rett syndrome). All of them had clear additional features (ID, epilepsy, muscular dystrophy, and MPAs) suggesting syndromic autism. One LP mutation occurred in the *PTPN11* gene, previously reported in ClinVar in association with Noonan syndrome. Some phenotypic feature indicated the presence of this syndrome, but some classical feature were missing (short stature and classical webbed neck). Segregation analysis was not possible in this family, however autosomal dominant inheritance

couldn’t be excluded. We detected two LP variants in the *RELN* gene with different phenotypes. A heterozygous *RELN* variant was present in a patient with epilepsy and multiple minor anomalies. This variant is also reported as pathogenic in the ClinVar database. Segregation analysis proved that the variant is inherited from the father, who had no epilepsy, however, incomplete penetrance is possible according to the literature (OMIM:616436). At the other case, a *de novo* variant occurred in the *RELN* gene. The lissencephaly phenotype is associated with autosomal recessive inheritance according to the databases, but we classified this variant as likely pathogenic instead of a VUS, because of its proven *de novo* status. To rule out other possible genetic causes, an additional commercial lissencephaly panel testing was also which gave the same result. Two variants were classified as VUS based on ACMG criteria. The *SPAST* variant was previously reported in ClinVar as LP, however, neither the patient nor the mother, who is also a carrier of the variant, have spasticity. In the last case, an unusual molecular event occurred. Two single nucleotide variants in cis position affected a single codon in the *AUTS2* gene,

resulting in p.E814M amino acid change. The phenotype is consistent with the literature (OMIM:615834) however, segregation analysis was not possible in the family. This patient died later as a consequence of a severe epileptic seizure.

MPAs were present in 84% of the cohort. In general, molecularly diagnosed syndromic patients had on average no more MPAs compared to non-syndromic cases (average number of minor malformations were 4.9/person in syndromic, and 5.03/person in non-syndromic cases, T-test p = 0.91). Strongest indicators of molecularly provable syndromic autism were ID, epilepsy or other plus neurological signs, or a specific constellation of MPAs as in the case of Fragile-X (FXS) and CHARGE-syndrome. Positive family history for psychiatric or neurological disorders were common in first and second-degree relatives. ASD occurred at a first degree relative at 14 patients (8%), and at a second degree relative at 17 patients (9,7%). Besides psychiatric disorders, epilepsy, muscle hypotony, speech development delay, intellectual disability (ID), dysmorphic features were also common.

Number of rare variants in the candidate gene panels

The total number of different rare variants was 370. Among the 101 candidate gene, 80 genes contained rare variants, and 44 genes contained a rare predicted pathogenic variant (CADD score ≥ 20 OR SVM = damaging). LoF mutations occurred in 8 genes: *AUTS2, CHD7, DHCR7, DMD, GNA14, MECP2, SHANK2,* and *SHANK3* gene. Among these genes, *AUTS2, CHD7, SHANK2,* and *SHANK3* can be considered highly intolerant to functional variants (RVIS < 5 percentile), which suggest a pathogenic role for these genes. We have used the software SORVA to assess whether the total number of rare missense and LoF variants is greater than expected in the tested genes. Significantly greater than expected number of rare variants were detected in 5 of the 101 tested ASD-linked genes: *AUTS2, NHS, NSD1, SLC9A9,* and *VPS13B.* Rare variant burden did not differ between females and males. There was 67 individual (38.5%) in the cohort who did not carry any rare, predicted damaging variant in the candidate genes, and many of the patients carried more than one rare variants.

Correlation of rare variant burden with ASD severity

Autism severity was given via total raw ADOS score, and ADOS-CSS. For the purpose of correlation analysis, we found that the CSS was not suitable, because most of the patients, who had ADOS fall into the most severe categories (8–10 score). ADOS raw total scores were better distributed among individuals, however, there was no correlation between ADOS raw total scores and rare variant burden by linear regression analysis. There was no correlation between rare variant burden and the number of MPAs neither.

Cluster analysis and gene enrichment in the phenotypic clusters

A total 22 feature were involved in the identified four clusters. The most frequent characterizing variables were present in more than one cluster, thus a significant overlap exists between the clusters. However, there are features, that particularly characterize certain clusters, e.g. a Cluster characterized by severe social disturbances, with prominent speech disturbance, or a Cluster probably fitting into the ASD+ADHD category. Gene enrichment analysis did not show

a significant overrepresentation of single genes in certain clusters. Syndromic cases were not enriched into a single cluster.

Mitochondrial dysfunction in ASD

MtDNA deletions were confrmed in 16.6% (10/60) of patients with ASD (mtdel-ASD). In 80% of this mtdel-ASD children we found rare SNVs in ASD-associated genes. In the IG panel no cause for primary MD was detected. Mitochondrial dysfunction co-occurred in both syndromic and non-syndromic ASD cases, it was considered mostly secondary due to mutations in non-OXPHOS genes or due to some environmental factors, such as CMV infection. There were some differences between ASD cases with and without mtDNA deletion regarding the clinical phenotype. Developmental regression, muscle hypotonia, and additional neurological signs were most common in the mtdel-ASD cases. Multisystemic abnormality appeared also more frequently.

CONCLUSION

In this study, we performed an analysis of rare single nucleotide and small INDEL variants in a Hungarian ASD cohort, detected by NGS panel testing, in order to identify syndromic autism cases and to assess the contribution of rare variants in formerly established ASD genes on a cohort level. Our study indicates that NGS panel gene sequencing can be useful in dedicated cases, where the clinical picture suggests a clinically defined syndromic autism (i.e., associated ID, epilepsy, neurological signs, a certain pattern of somatic malformation, or positive family history). The necessity of unselected NGS panel screening in the clinic remains controversial, because of uncertain clinical utility, and difficulties of the variant interpretation. The detected rare variants may still significantly influence autism risk and subphenotypes in a polygenic model. To detect the effects of these variants large cohorts are needed. As knowledge will increase about the contribution of these rare variants on the phenotype, an individual assessment might also be

beneficial in the future for personalized management of patients with ASD.

Mitochondrial dysfunction is one of the most common associated abnormality in ASD. MtDNA deletions are not isolated genetic alterations found in ASD; they coexist with other ASD-associated genetic risk factors. Many clinical features as well as genetic and biochemical abnormalities associated with ASD could be related to mitochondrial dysfunction. The occurring mtDNA deletions in ASD may be mostly a consequence secondary mitochondrial damage, caused by the alterations of non-OXPHOS genes, or because of the harmful effect of environmental factors.

BIBLIOGRAPHY OF THE CANDIDATE'S PUBLICATIONS

Publications related to the PhD thesis

Balicza P*, Varga NA*, Bolgar B, Pentelenyi K, Bencsik R, Gal A, Gezsi A, Prekop Cs, Molnar V, Molnar MJ. (2019) Comprehensive analysis of rare variants of 101 autism-linked genes in a Hungarian cohort of autism spectrum disorder patients. Front. Genet. 10:434-15. **IF: 3,517**

Varga NA, Pentelényi K, Balicza P, Gézsi A, Reményi V, Hársfalvi V, Bencsik R, Illés A, Prekop C, Molnár MJ. (2018) Mitochondrial dysfunction and autism: Comprehensive genetic analyses of children with autism and mtDNA deletion. Behav Brain Funct. 14:4-14. **IF: 2,457**

Kékesi A*, Varga NÁ*, Molnár MJ. (2014) Az autizmus spektrum betegség genetikája Gyermekgyógyászat. 65:327-332.

Publications not related to the PhD thesis

Varga NÁ, Barsi P, Molnár MJ. (2015) A Leigh-szindróma a genetikailag legheterogénebb mitokondriális betegség. Gyermekgy. Tov. Szeml. 20:75-77,3p.

Buza K., Varga NÁ (2016) ParkinsoNET: Estimation of UPDRS Score Using Hubness-Aware Feedforward Neural Networks. *App. Art. Intell., 30*(6), 541–555. **IF: 0,652**

Tamás G, Kovács N, Varga NÁ, Barsi P, Erőss L, Molnár MJ ,Balás I. (2016) Deep brain stimutation or thalamotomy in fragile X-assiciate tremor/ataxia syndrome? Case report. *Neurologia i Neurochirurgia Polska, 50*(4), 303–308. **IF:0,857**

Balicza P, Terebessy A, Grosz Z, Varga NA, Gal A, Fekete BA, Molnar MJ. (2018) Implementation of personalized medicine in Central-Eastern Europe: pitfalls and potentials based on citizen's attitude. *EPMA Journal*, *9*(1), 103–112. **IF:3,900**