CLINICAL APPLICATION OF NEXT GENERATION SEQUENCING METHODOLOGY IN THE DIAGNOSIS OF RARE NEUROLOGICAL DISEASES

PhD thesis

Péter Balicza MD

János Szentágothai Doctoral School of Neurosciences

Semmelweis University

Supervisor: Mária Judit Molnár, MD, D.Sc

Official reviewers: Zoltán Wiener, Ph.D
Ágnes Sebők, MD, Ph.D

Head of the Final Examination Committee: János Réthelyi, MD, Ph.D

Members of the Final Examination Committee: Ildikó Szatmári, Ph.D
Zoltán Liptai, MD, Ph.D

Budapest
2018

INTRODUCTION

Paradigm shift in the diagnosis and classification of neurodegenerative diseases

Presently about 8000 rare diseases are known, 80% of which are genetic in origin, and a significant proportion (>50%) is associated with neurological symptoms. Neurogenetic diseases encompass the whole spectrum of neurological symptomatology. They can present as dementia, muscle dystrophy, movement disorder, spastic paresis, epilepsy, neuropathy, and involve the rare monogenic forms of the more frequent neurological diseases. The molecular diagnosis is difficult based on clinical findings, because variants in different genes can be the cause of very similar phenotypes (so called locus heterogeneity). On the other side, different variants in a certain gene may result in different phenotype (so called pleiotropy). Because of the difficulty and cost of diagnostics, many times people suffering from neurogenetic diseases never get a molecular diagnosis.

The next generation sequencing (NGS) is increasingly utilized last years in the clinical genetic research and diagnostic, which resulted in a paradigm shift. The essence of next generation sequencing methodology is the ability of massive parallel sequencing, which results in faster and more cost effective base sequence determination, compared to the Sanger methodology. During the NGS process large amount of data is generated. The bioinformatic processing and interpretation of the results is often taking more time than the laboratory process. Exome sequencing generates about 20,000 variants, the panel sequencing a few hundred variants, from which we need to filter the one or two causative variant associating with the phenotype. The filtration method is usually based on inheritance pattern, evolutionary conservation, the effect of the variant on biochemical structure of the protein, minor allele frequency, information in databases. There are also software which help to filtrate variants based on symptomatology. The advantage of NGS studies is that they offer opportunity to discover novel gene-phenotype associations.

As the result of the NGS studies of last years, or view has changed significantly on the classification of neurodegenerative diseases. Diseases, formerly described as separate entity became connected by a common gene. Instead of discrete phenotypes, in many instances spectrum diseases were outlined. In the last years, we increasingly deem

ataxia-spasticity spectrum, and the different „axonopathies" (hereditary spastic paraparesis, ataxia, hereditary neuropathies) a unified group of diseases. In this model, the different gene defects cause a certain level of vulnerability in different neuroanatomical structures. Hereditary spastic paraparesis is associated with variants in about 70, amyotrophic lateral sclerosis 33, and ataxias more than 100 genes. Certain genes (e.g. *SPG11*, *SETX*, *SPG7*, and *BSCL2*) are shared among these phenotypes. For this reason NGS is suitable for the study of ataxia-spasticity spectrum diseases.

Ethical concerns raised by the clinical application of NGS

The increased amount of information produced by NGS technologies produce novel ethical challenges for clinical geneticist, of which incidental findings and variants with uncertain significance has to be highlighted. Because of the high throughput sequencing many variants may be detected, not related directly to the phenotype which is studied, but having significance for the tested person. Until now there is no consensus on how these findings should be reported and in what depth. Further difficulty arises with the variants of uncertain significance, detected frequently by NGS, of which interpretation is currently limited, but could potentially bear information for the tested individual. Further ethical questions are raised by the blurring line between diagnostic investigations and clinical research, and the sharing of NGS data. In order to draw up recommendations on the above questions we have to know the opinion of the Hungarian population, but studies in this field are limited in Hungary and Central-Eastern Europe.

OBJECTIVES

Our objectives were the followings:

1. Study the utility of next generation sequencing methodology in the clinical diagnostics (diagnostic rate of Sanger vs. next generation sequencing panel and whole exome sequencing) in the ataxia-spasticity spectrum diseases.

2. To perform genetic epidemiologic studies among patients diagnosed with hereditary spastic paraparesis.

3. Study the genotype-phenotype correlations in ataxia-spasticity spectrum diseases (identify HSP, ALS, ataxia phenocopies).

4. Identify new causative gene in a complex mitochondrial disease.

5. Study the attitude of Hungarian citizens towards high throughput sequencing, and discussing the ethical questions related to it.

METHODS

Patients involved in genetic studies

In the hereditary spastic paraparesis (HSP) study we examined altogether 100 proband (99 proband with HSP phenotype, and one proband with onset as ataxia, later proved to be HSP). The *SPAST* gene was Sanger sequenced in 34 proband. In 11 case, where *SPAST* and *ATL1* Sanger sequencing was negative, we performed whole exome sequencing. In two probands, where the diagnostic muscle biopsy showed mitochondrial changes, the whole *SPG7* gene was sequenced. At 44 probands NGS panel sequencing was performed without previous Sanger sequencing of HSP genes. At further 34 proband targeted Sanger sequencing of the second exon of *SPG7* gene was done, for the detection of the p.Leu78* variant. In an 18 years old male patient, who presented himself with cerebellar ataxia we performed trio exome sequencing. We performed panel sequencing at further 14 patients with motoneuron disease (13 patients with amyotrophic lateral sclerosis and one patient with primary lateral sclerosis). We performed trio exome sequencing in a family with a complex mitochondrial phenotype.

Methods of the genetic studies

DNA isolation was performed from blood samples with „QIAmp DNA blood kit". We have used ABI Prism 3500 DNA sequencing machine for Sanger sequencing. For NGS panel sequencing a custom Agilent Haloplex, and Agilent Sureselect QXT target enrichment kit was used, sequencing was performed on Illumina MiSeq platform. Whole exome sequencing in the HSP patients was done with Agilent Sureselect Human All Exon V4 target enrichment kit on Illumina HiSeq platform. At the family with complex mitochondrial phenotype NimbleGen SeqCap EZ Human Exome Library v3.0 was used for target enrichment. Sequencing was performed on Illumina HiScanSQ platform. At the patient with cerebellar ataxia we used Nextera Rapid Capture kit, and exome sequencing was done on Illumina HiSeq 2500 platform. For the analysis process we used the Agilent Surecall v2.1, GEM.app, Samtools, Illumina Basespace BWA Enrichment, SnpEff, SnpSift, Exomiser, HSF3, Alamut and VariantAnalyzer software.

Methods of the questionnaire study

We created a questionnaire for the study of attitude towards high throughput sequencing genetic tests. The questionnaire contained 37 questions, which was available on paper and online. We collected responses on paper questionnaire at our neurogenetic outpatient clinic from patients and patient relatives. The web link for the questionnaire was sent out in email for patients, who previously provided their email address for us. The web link was also made available on social media. Further responses were collected from medical students who took neurogenetic classes, and from further citizens who attended our annual open lecture. The statistical analysis involved chi-square test, binary logistic regression analysis, Kruskal-Wallis test, and Mann-Whitney U test.

RESULTS

Utility of next generation sequencing in the diagnostic of spasticity-ataxia spectrum diseases

The targeted Sanger sequencing of the *SPAST* gene identified the molecular background of the disease in 20,6%. With exome sequencing among *SPAST* and *ATL1* negative patients the pathogenic variant was found in 36,4%. The diagnostic rate of panel sequencing (without previous *SPAST* and *ATL1* sequencing) was 29,5%. The NGS methodology used by us resulted in an overall diagnostic rate similar to other studies worldwide.

The molecular epidemiology of HSP in Hungary

Molecular background of the disease was identified in 26,3% (26 proband) of the whole HSP cohort. Pathogenic variants were most frequently detected in the *SPAST* gene (12 probands), followed by pathogenic variants in *SPG7* (six probands), *SPG11* (two probands), *ATL1* (one proband), and *NIPA1* (one proband). At 11 patients the detected variant was previously not reported in the literature. At four patients, originally diagnosed with HSP, phenocopies was proved. At two probands adrenoleukodystrophy (*ABCD1* variant), at one proband Krabbe disease (*GALC* variants), and at one patient FAHN (*FA2H* variants) was diagnosed.

Genotype phenotype correlations in HSP

With our studies we have broadened the available literature about phenotype genotype associations in spasticity-ataxia spectrum diseases. We observed significant intra- and interfamilial variability in the clinical symptoms associated with *SPAST* variants. Certain variant associated with cHSP. The *SPAST* p.Pro435Leu variant associated with ataxia, microcephaly, and white matter lesion. The *SPAST* p.Ala428Pro variant associated with early onset depression. In a family with *SPAST* p.Thr614Asnfs*16 variant in the affected family members, ataxia associated in two patients. Patients with

splice variants had pHSP, at the case of *SPAST* c.1536+1G>A variant we suspected incomplete penetrance.

We diagnosed SPG7 in six probands, at five probands the *SPG7* p.Leu78* variant was present (at four proband as a part of homozygous our compound heterozygous variant). At one proband it was detected as a simplex heterozygous variant suggesting that it occur as autosomal dominant. Complicating symptoms were present in two probands. At one case distal neuropathy, at the other case optic atrophy associated. The possibility of autosomal dominant inheritance was also raised with the *SPG7* p.Gly344Asp variant. The patient harboring this mutation as a part of a compound heterozygous variant was diagnosed with cHSP. The associating features were ataxia and cognitive decline.

Our findings supported the hypothesis, that the variant *SPAST* p.Ser44Leu and SPG7 p.Ala510Val has modifier effect on the phenotype.

SPG11 was diagnosed in three probands. At two probands, who were examined in the HSP cohort, the symptoms correlated closely with the phenotype described in the literature. In one case investigated with cerebellar ataxia the exome sequencing found compound heterozygous variants in the *SPG11* gene.

We observed phenotypic overlap between cHSP and leukodystrophies with variants in the *ABCD1*, *GALC*, and *FA2H* genes. The brain MRI was initially negative in the patients with adrenomyeloneuropathy (*ABCD1* variant), and the phenotype was similar to HSP. At the patient with Krabbe disease (*GALC* variants) the brain MRI showed changes consistent with leukodystrophy, but the phenotype was pure HSP. The variants of the *FA2H* gene are known to associate with HSP (SPG35) and leukodystrophy (FAHN disease). With the variants detected by us diffuse white matter lesions and complex neurologic symptoms associated.

HSP panel sequencing in amyotrophic lateral sclerosis and primary lateral sclerosis

We detected rare, probably protein structure changing variants in 50% of the examined cohort (7/14 proband) in genes associated with neurodegenerative diseases. Four probands (28,5%) had variants in more than one gene simultaneously. We identified rare variants in genes associated with familiar ALS, in the classic HSP genes, and also

in nuclear mitochondrial genes, which represent a part of the HSP differential diagnostic. The findings fit into the oligogenic model of ALS. In three patients rare *POLG* variants were detected. At one patient compound heterozygous, previously described pathogenic variants were present. At one patient possibility of seipinopathy (*BSCL2* variant) was raised.

Identification of the *MSTO1* gene in the background of a human disease

In a family with a complex mitochondrial phenotype the exome sequencing identified the *MSTO1* gene, which was not previously linked to any human disease. At the proband myopathy, ataxia, psychiatric symptoms, minor malformations, endocrine disorders were present. At the offspring of the proband psychiatric, endocrine symptoms, minor malformations, neurological abnormalities were present with various severity. We suspected mitochondrial disease based on the complex phenotype, the result of the muscle biopsy and lactate stress test at the proband, but no mtDNA alteration was detected. The exome sequencing detected a pathogenic heterozygous variant in the *MSTO1* gene (p.Val8Met), which segregated with the phenotype in the family. The exact function of the MSTO1 protein is only partially determined, but the functional studies performed by the member of our workgroup (Dr. Anikó Gál) confirmed, that the pathogenic variants of the *MSTO1* gene cause mitochondrial pathology.

The attitude toward high throughput genetic tests in Hungary

Our questionnaire was filled out by 657 person. The respondents were interested in genetic tests with medical utility, and also for incidental findings, and variants with uncertain significance. Respondents with higher perceived genetic influence score showed higher interest for the whole genome sequencing and sequencing for risk modifier genetic variations. Respondents with higher self-rated genetic familiarity score had more often a high perceived genetic influence score. Based on the bivariate analyses interest was greater at male respondents, at respondents not working in the healthcare system, and with the possibility of consultation with a physician. Attitude toward direct

to consumer genetic test was most strongly affected by the age, and the sex of the respondent, based on the binary logistic regression analysis.

CONCLUSIONS

Based on our research we concluded that both our custom NGS panel, and whole exome sequencing is effective in the molecular diagnostics of spasticity-ataxia spectrum diseases. In 26,3% of the whole HSP cohort we determined the molecular background of the disease, and performed comprehensive genetic testing in a HSP cohort first in Hungary. Based on the findings we formulated recommendations for the health care system regarding cost effective diagnostic of Hungarian HSP patients. We found that SPG7 variants are much more frequent in Hungarian HSP patients as it was previously assumed. We have broadened the available genotype-phenotype correlations in the literature on variations in the *SPAST, SPG7,* and *SPG11* genes. With the panel, and exome sequencing we identified phenocopies in our HSP cohort, and detected HSP associated gene (*SPG11*) variation as a cause of cerebellar tremor in a patient. We have detected heterozygous rare variants in HSP-associated genes, and in the *POLG* and *BSCL2* gene in ALS patients. In a family with a complex mitochondrial phenotype, we identified a new gene (*MSTO1*), which was previously not linked to human disease. Based on our questionnaire we concluded, that in Hungary there is a need for high throughput genetic tests, but also for genetic counseling. Profession, age, sex, self-rated genetic familiarity, and perceived genetic influence affect the attitude toward genetic testing, which is important information regarding realization of personalized medicine, and genetic counseling.

BIBLIOGRAPHY OF THE CANDIDATE'S PUBLICATIONS

Publications related to the PhD thesis

1.) Balicza P, Grosz Z, Gonzalez MA, Bencsik R, Pentelenyi K, Gal A, Varga E, Klivenyi P, Koller J, Züchner S, Molnar MJ. (2016) Genetic background of the hereditary spastic paraparesis phenotypes in Hungary - An analysis of 58 probands. J Neurol Sci, 364:116-121. IF: **2,295**

2.) Gal A, Balicza P, Weaver D, Naghdi S, Joseph SK, Várnai P, Gyuris T, Horváth A, Nagy L, Seifert EL, Molnar MJ, Hajnóczky G. (2017) MSTO1 is a cytoplasmic pro-mitochondrial fusion protein, whose mutation induces myopathy and ataxia in humans. EMBO Mol Med, 9: 967-984. **IF: 9,249**

Publications not related to the PhD thesis

1.) Bock I, Németh K, Pentelényi K, Balicza P, Balázs A, Molnár MJ, Román V, Nagy J, Lévay G, Kobolák J, Dinnyés A. (2016) Targeted next generation sequencing of a panel of autism-related genes identifies an EHMT1 mutation in a Kleefstra syndrome patient with autism and normal intellectual performance. Gene, 595: 131-141. **IF: 2,415**

2.) Szatmari B, Balicza P, Nemeth G, Molnár MJ. (2017) The Panomics Approach in Neurodegenerative Disorders. Curr Med Chem, epub ahead of print. IF: -

3.) Balicza P, Bereznai B, Takáts A, Klivényi P, Dibó G, Hidasi E, Balogh I, Molnár MJ. Az LRRK2 gyakori G2019S-mutációjának hiánya 120, korai kezdetű magyar Parkinson-beteg esetében. (2012) Ideggyogy Sz, 65: 239-242. **IF: 0,348**

4.) Györfi O, Moarcas M, Balicza P, Macerollo A, Struhal W, Sellner J. (2016) European junior neurologists perceive various shortcomings in current residency curricula. Acta Neurol Scand, 134:232-237. **IF: 3,087**