THE APPLICATION OF THE NEXT GENERATION SEQUENCING IN THE DIAGNOSIS OF RARE GENETIC DISORDERS

PhD. Thesis

GÁBOR KOVÁCS

SZEGED

2016

THE APPLICATION OF THE NEXT GENERATION SEQUENCING IN THE DIAGNOSIS OF RARE GENETIC DISORDERS

PhD. Thesis

Gábor Kovács

Supervisors

Dr. Emőke Endreffy, Dr. Zoltán Maróti

Department of Paediatrics

Faculty of Medicine

University of Szeged

SZEGED

2016

Introduction

Rare diseases (RDs), also called orphan diseases are those which affect only a very small portion of the population and the majority of them (~80%) are genetic. According to the latest international agreement, the prevalence threshold for RD not more than 5 affected persons per 10,000 (European Parliament and of the Council of 16 December 1999). Up to date more than 6000 RDs are described by a broad diversity of symptoms. Many of them are characterised by mental retardation and affects 50-70% of children and even today around 30% of the affected patients will die under the age of five. The common symptoms can mask the basic rare diseases which can lead to misdiagnosis and delays treatment. In 1997 the Orphanet database was created which aim is to provide information about rare disease to clinicians and patients in order to contribute to the successful diagnosis and to better treatment. RDs were registered with one offered name and as many synonyms with an ORPHA number (www.orpha.net). As the result of this it makes even more difficult to even guess the exact number of rare diseases. Majority of RDs are monogenic, however due to the „umbrella terms" used for disease groups with overlapping phenotype there are numerous complex diseases that are caused by few to hundreds of genes (nephrosis, schizophrenia, ectodermal dysplasia, chronic obstructive pulmonary disease, etc.).

"*The average patient with rare disease visits 8 physicians and receives 2 to 3 misdiagnosis before being correctly diagnosed. On average, this takes 7-6 years."*

/Shire HGT: Rare Disease Impact Report-2013/

Up to recently the clinical genetic diagnosis of complex diseases were only partially solved. Sanger sequencing of hundreds of genes is not feasible in diagnostics due to the enormous manual work and costs of the analysis. This challenge was partially handled by Orphanet where the most frequent genes or specific mutations causing various diseases were analysed in specific expert centres. Our genetic diagnostic laboratory is an expert centre of several RDs, including monogenic and complex diseases. Among others, we use diagnostic methods to analyse the most frequent X linked type (COL4A5 gene) Alport syndrome and X linked type ectodermal dysplasia (EDA gene). In many cases the traditional (linkage analysis, high resolution melting analysis, Sanger sequencing based) methods investigating only the most frequent genes does not reveal the genetic cause of the disease.

Recently with the advancement of high-throughput massively parallel sequencing technologies (new generation sequencing, NGS) it became possible to sequence more genes and more samples at the same time. Since we had several unsolved cases of complex disease

we wanted to establish and test the feasibility of NGS. In my thesis I chose to work with Alport syndrome and ectodermal dysplasia in order to improve their current molecular diagnostic approach in our laboratory, thus my thesis has two parts:

Alport related disorders

Familiar benign haematuria (FBH) and Alport syndrome (AS) are familial haematuria diseases which in case of AS regularly escalate to chronic kidney disease (CKD) stage 5 (formerly referred as end stage renal disease). AS patients usually have sensorineural high-tone deafness and ocular abnormalities affecting the lens and fundus. Today more focus has been placed on treating patients early to prevent or delay future end stage kidney damage. Although, the pathogenesis of CKD is multifactorial, some of the suggested therapeutic interventions (anti-hypertensive therapy, glycaemic control, anti-proteinuric therapy, renoprotection, and life style management such as restricted protein intake, cessation of cigarette smoking and chronic analgesic-abuse) are encouraging. These preventive steps the more earlier are implemented the more efficient they are.

Sequential (one by one) genetic testing for mutations in *COL4A3-A4-A5* genes has become an integral part of the clinical evaluation. Since all three genes are large (contains 52, 48 and 51 exons, respectively) the use of conventional Sanger sequencing is time-consuming, expensive and can suffer from some technical limitation (such as failing to detect insertion/deletion with certain sizes in a heterozygote subject). One way to overcome these problems is to sequence all three genes simultaneously using targeted NGS sequencing.

Ectododermal dysplasia related disorders

Ectodermal dysplasias (ED) are inherited disorders characterized by abnormal morphogenesis of ectoderm. Ectoderm is one of the three primary germ layers during the development of early embryo. Ectodermal dysplasia is an umbrella term representing hereditary diseases of a large group of conditions in which two or more ectodermally-derived tissues fail to develop. There are more than 150 types of Ectodermal dysplasia described so far, with a combined incidence of 7/10000. Based on the inheritance there are autosomal recessive (AR), autosomal dominant (AD), X-linked recessive (XR) and X-linked dominant (XD) forms. Because of the high number of the possible causative genes and the observed different phenotypes it is difficult to establish the proper diagnosis for the clinicians and geneticists.

The following tissues could be affected in ED:

- hair: the hair is absent on the scalp, sparse and light with abnormal texture. The eyebrows and eyelashes or the other body hair may absent or sparse.
- nails: nail usually poorly developed both on the fingers and on the toes. Usually thick, sharp, stands out and fragile.
- teeth: may be missing or reduced in number. Teeth are widely spaced and tapered/peg shaped.
- sweat glands: missing or scattered. Due to their functional problems the body is unable to maintain the normal body temperature. As a result extreme high body temperature is frequent and the patients are unable to tolerate high outside temperature.

Aims

1. Investigation of the underlying causative genetic alterations by next generation sequencing in 14 COL IV unsolved nephropathies families and in 3 unsolved nephropathies individuals.
2. Try to give a better description to the controversial inheritance (AR, AD) mode of Alport and TBMN
3. To identify disease causing variants by clinical exom sequencing in 3 families affected with different types of ectodermal dysplasia.

Patients and methods

Alport related disorders

Our Hungarian cohort consisted of 3 individuals (where family members were not available for either clinical or genetic analyses) and 14 COL IV nephropathy families where clinical data and inheritance of clinical symptoms supported Alport/FBH diagnosis and other possible clinical causes were excluded. All volunteer family members underwent urinalysis and renal function evaluation. Kidney biopsy was not performed on all proband. Criterias for FBH were: mild persistent haematuria (25-30 RBC/HPF), mild (0.5-1.0 g/day) or no proteinuria and first appearance after 10 years of age. Criterias for Alport were: persistent haematuria (>30, 50-100 RBC/HPF) and persistent proteinuria (0.5-5.0 g/day). After obtaining written informed consent peripheral blood samples were collected from both the affected and unaffected

members of the families. We used anonymised DNA samples from our licensed biobank as the 66 Hungarian non AS/FBH control samples.

We designed a custom AmpliSeq panel for the simultaneous analysis of COL4A3, A4 and A5 genes that contained all the coding and flanking regions of the analysed genes. Sequencing was performed on IonTorrent PGM platform. Variant calling was done by TorrentSuite 4.2 software.

Ectodermal dysplasia related disorders

We used clinical exom sequencing to analyse 3 families affected with different types of ectodermal dysplasia.

Family I: 35 years ago a one year old boy had been diagnosed with ectodermal dysplasia. Young man in the process of family planning turned up at our clinic to ask for genetic counselling. His symptoms (fever periods, hypodontia, lack of hair) indicated hypohidrotic ectodermal dysplasia. We investigated his affected mother, grandmother and 3 unaffected family members.

Family II: Four years ago a 14 years old boy was treated at our clinic with poisoning. During his clinical investigation hypohidrotic ectodermal dysplasia (HED) diagnosis was established. The patient's reduced sweating, sparse hair and malformed teeth and the absence of eyebrows, eyelashes has suggested HED. Additional clinical observations such as the syndactyly on both feet (between second and third toes) and the double nipples on the right side of his chest also supported genetic syndrome. During the clinical investigation of the family we found other affected members with milder symptoms (only tooth agenesis). It seemed that two different phenotypes segregate in the family which indicate that might not only one gene responsible for all observed clinical symptoms.

Family III: 5 years old boy with Indian origin was born from a 3rd degree consanguineous marriage. The child has absent hair, eyebrows, eyelashes, severely dystrophic nails and syndactyly between the second and third toes. His dental aspects and intelligence were normal. In the familial anamnesis we observed the dystrophic nails and mild proximal syndactyly in cases of the affected boy, his mother and his father.

We chose one affected member from every family to analyse by NGS. To sequence all 141 genes which known to cause ECD and its related disorders we used TruSight One clinical exom sequencing (CES) panel from Illumina. The exome sequencing is a targeted sequencing approach which is restricted to the protein-coding regions of the genome (exom). This panel

contains 4813 genes which are confirmed to cause monogenic illnesses (HGMD, OMIM). The sequencing was performed on Illumina MiSeq. Variant calling was done by BWA + GATK HaplotypeCaller softwares.

Validation by Sanger sequencing

The presence of the identified variations in all ECD and AP/TBMN samples were validated by Sanger-sequencing. We performed the segregation analysis of the putative pathogenic variants in the affected and the unaffected family members. Using this approach we could confirm the pathogen status of the variants.

Results

We present a new, efficient, amplicon based NGS protocol for simultaneous analysis of the COL4A3, A4 and A5 genes.

We confirmed the pathogenic status of different alterations in 80 samples (20 probands + 60 family members). We found 14 AS/TBMN unpublished mutations and 4 new, previously unpublished ECD mutations

Summary and conclusions

1. We successfully employed a cost-effective (< 500 EUR/sample) NGS based method (Custom AmpliSeq panel, bioinformatic evaluation) for simultaneous investigation of COL4A3-A4-A5 genes. Our approach proven to be suitable for the detection of SNPs and smaller INDELs.
2. We found the causative mutations in all the investigated 17 COL4 nephropathy cases.
3. We have successfully identified and validated pathogenic alterations (4 non-published mutation) by clinical exom sequencing in the sample of ECD patients. The clinical diagnosis was confirmed by our genetic results.
4. Based on the identified mutations and confirmed genetic status positive family planning and proper therapy can be offered to the analysed families.
5. Based on our results we think:
    - Alport and its related diseases are part of a spectrum disorder, where the biological base of the different phenotypes is the amount of fully functional (or impaired/null) Type IV collagen triple-helix structure.
    - the distinct AD inheritance type of Alport syndrome is likely the result of technical error or the differences in the interpretation of the clinical symptoms
    - the targeted NGS can be a reliable method for the genetic diagnosis of disorders which are clinically well identified and caused by only few genes
    - clinical exome sequencing proved to be suitable for the genetic analysis and diagnostic of complex diseases

Acknowledgements

There are number of people without whom this thesis might not have been written, and to whom I am greatly indebted:

To my supervisor **Dr. Emőke Endreffy** who supervised my scientific work when I was an undergraduate student and let me continue my work as a PhD student.

Let me express my gratitude to **Dr. Zoltán Maróti** for his leaderships and professional advice which I have had the good fortune to benefit from while finalizing my thesis and ever since. I thank to **Dr. Tibor Kalmár** for his leaderships during my PhD studies. I thank you both again for your trust and friendship which allows me to stand here in front of you this day.

I should also like to thank *Dr. Zsolt Rázga* for his kind help during my PhD study.

I am also very grateful for my dear colleagues: *Dr.Ibolya Haszon, Dr. Eszter Karg, Dr. Mária Sinkó, Dr. Eszter Karg* and *Dr. Ákos Baráth* whose suggestions helped to improve the quality of my work,

I would like to say a special thanks to *Mrs Erzsebet Borzási, Mrs Ildikó Csipő, Mrs Mária Mustoha* and *Mrs Tünde Vida-Szűcs* for the technical assistance.

To the ex-chair of the Department of Pediatrics *Prof. Dr. Sándor Túri* and to the present chairman of Department of Pediatrics *Dr. Csaba Bereczki* for making my research possible and being so supportive during my undergraduate and doctoral studies.

I would like to thank my family for their unconditional support, both financially and emotionally throughout my studies.

Publications

Publications of this thesis

Gábor Kovács, Tibor Kalmár, Emőke Endreffy, Zoltán Ondrik, Béla Iványi, Csaba Rikker, Ibolya Haszon, Sándor Túri, Mária Sinkó, Csaba Bereczki, Zoltán Maróti. Efficient Targeted Next Generation Sequencing-based workflow for differential diagnosis of Alport-related disorders; *PlosOne*; 2016 DOI: 10.1371/journal.pone.0149241; [IF: 3,2]

Endreffy Emőke; **Kovács Gábor**; Horváth Emese; Karg Eszter; László Aranka; Túri Sándor. Ectodermalis Dysplasias férfi molekuláris genetikai vizsgálata. *Gyermekógyászat*; 2008; 2008/59. p. 85-88.

Additional related publications

Zsolt Razga, **Gabor Kovacs** et al. Upregulation of the L-type Calcium Channel in Renin-positive Smooth Muscle Cells of Arterioles in the Kidneys of Rats with STZ-induced Diabetes; *Analytical and Quantitative Cytology and Histology;* 2015; 37(4):214-20. [IF: 0,5]

Zsolt Razga, **Gabor Kovacs**, Nikolett Bódi, Petra Talapka, and Jens Randel Nyengaard. Heterogeneous Downregulation of Angiotensin II AT1-A and AT1-B Receptors inArterioles in STZ-Induced Diabetic Rat Kidneys. *BioMed Nephrology*; 2014; ID 947506. [IF:2,8]