Identification of incompletely penetrant variants and interallelic interactions in autosomal recessive disorders by a population genetic approach

PhD thesis

Ágnes Mikó

Doctoral School of Clinical Medicine
Semmelweis University

Supervisor: Kálmán Tory, Ph.D

Official reviewers: Zsófia Nemoda, Ph.D
Balázs Egyed, Ph.D

Head of the Final Exam Committee: Darvas Zsuzsanna, Ph.D

Members of the Final Exam committee: Haltrich Irén, Ph.D
Balogh István, Ph.D

Budapest

2022

1. Introduction

Among differently inherited genetic diseases, a specific phenotype is most closely determined by the genotype in case of autosomal recessive inheritance. Autosomal recessive (AR) diseases result from loss of function, incomplete penetrance occurs only in exceptional cases. Penetrance expresses in which proportion individuals develop a disease, carrying the same genotype. Incomplete penetrance means, if there are individuals, who carry the same allele, but do not develop the specific disease phenotype.
Determining the frequency of certain variants in large populations offers a new opportunity to define the penetrance of pathogenic variants.
In autosomal recessive diseases penetrance is affected by several factors; the type of the pathogenic variant (it happens that the variant does not lead to the development of a disease due to its hypomorphic nature), epigenetic factors, di-and oligogenic inheritance, modifying genes and interallelic interactions.

Current research based on a previous work of our research team, which identified by population genetic approach, the first interallelic interaction that results in incomplete penetrance.
This variant is a common variant (R229Q, MAF: 3,6 %, rs61747728) of the *NPHS2* gene, the mutations of which are commonly responsible for steroid-resistant nephrotic syndrome. Steroid-resistant nephrotic syndrome (SRNS) is responsible for 6-7% of chronic renal failure cases during childhood. The two main causes of SRNS are monogenic podocytopathies and immune dysfunction. Podocytopathies result from mutations of genes encoding proteins of the glomerular filtration barrier (*NPHS2, WT1, NPHS1, MYO1E, TRPC6, INF2*). The most commonly affected genes encode proteins

implicated in podocyte cell adhesion, structure, differentiation and function.

Biallelic *NPHS2* mutations typically cause end-stage renal disease up to the age of 10 years. Podocin, encoded by *NPHS2*, is exclusively expressed in podocytes. As a member of the stomatin superfamily, it is membrane-anchored. The function of podocin is however not entirely understood.

This mutation-dependent recessive inheritance in human autosomal recessive diseases demonstrated by our research team was previously unknown.

The observation was based on the fact that this common allele of NPHS2, which showed a clear enrichment in patients treated for genetic SRNS, did not occur in homozygous form in patients, but only in association with certain mutations in the last two (7-8) exons of the gene.

The hypothesis of interallelic interactions at a molecular level was previously understood by our research team on the basis of molecular dynamic measurements performed by chemists. Later, Förster resonance energy transfer (FRET) allowed us to indirectly examine the structure of podocin heterooligomers.

Based on our measurements we learned that podocin oligomerises exclusively through the C-terminal region and based on molecular modelling studies, we saw that the R229Q makes the C-terminal helical tail less flexible, which makes it prone to form abnormal heterodimers with some specific C-terminal missense mutants also affecting the dimerization.

These findings have a fundamental impact on the diagnosis, treatment and genetic counselling of *NPHS2*-related SRNS. If one parent carries a mutation in *NPHS2* with no dominant negative effect, and the other is heterozygous for R229Q, their child is not at risk of developing SRNS.

The penetrance of incompletely penetrant (IP) variants is difficult to estimate and interallelic interactions may easily go undetected in the general population.

We, therefore, aimed to identify novel IP variants and interallelic interactions.

2. Aims

- Similarly to the example of the R229Q variant, identification of incompletely penetrant variants in common autosomal recessive diseases by population genetic approach.
  Identification of interallelic interaction dependents among incompletely penetrating variants.

- Our goal was to develop a computerized algorithm written in R program, thus, to provide a way to a quick and efficient analysis of the growing literature.

- Proving experimentally the changed structure of pathogenic podocin hetereooligomers.

- Establishment of a clinical guideline for the assessment of unknown pathogenic associations of the *NPHS2* R229Q variant.

3. Methods

Examination of variants by population genetic method

We have created a population genetic algorithm that gives the opportunity to calculate the variants allele frequency. Due to databases in the medical literature, which corresponds to a control population group, with no monogenic disorders, we had to create a patient population to apply the population genetic method.

Selection of autosomal recessive disorders

Genes implicated in AR disorders (n = 1981) were downloaded from OMIM (Online Mendelian Inheritance in Man). As a population-genetic approach is primarily informative in frequent disorders with a large variety of variants, we chose four frequency criteria to select the genes of interest: (1) the prevalence of the associated disorder in the literature is estimated to be >1:200,000; (2) >100 different pathogenic variants are reported in the Human Gene Mutation Database (HGMD); (3) the cumulative AF of loss-of-function (LOF) variants in the European non-Finnish population is >0.02% according to gnomAD and (4) the number of published European patients in the medical literature >100; (5) previously reported interallelic interactions in in vitro experiments. Based on these criteria 17/1981 genes have been selected: *ASL, ATP7B, CAPN3, CFTR, CTNS, DHCR7, GAA, GALNS, GALT, IDUA, MUT, NPHS1, NPHS2, PAH, PKHD1, PMM2, SLC26A4.*

Collection of patient data

The data of the patients were collected manually from the literature cited in the HGMD database, using the Google-scholar and Pubmed search engines.
Only cohort studies sequencing all the coding exons have been included (n = 341), and those patients who carried variants on both alleles. The ethnic origin of patients was classified according to gnomAD's database thus, for example the European group was divided into non-Finnish and Finnish groups. From 438 publications, we collected data from 12,048 patients suffered in one of the 17 diseases of choice and who carried a total of 3296 different variants.

Synchronizing the variant nomenclature

For each gene, the canonical transcript found in the GRCh37.p13 (hg19) human reference assembly was selected according to the definition of Ensembl. The nomenclature of each variant was validated in the patient database according to the Human Genome Variation Society (HGVS) standard using Mutalyzer Name Checker.

Classification of the variants

Variants were classified according to three criteria: (1) LOF/missense, (2) domain localization, and (3) null/hypomorphic. The biological classification (LOF/missense) and the domain localization were used to assess the interallelic interactions. The clinical (null/hypomorphic) and biological classifications also served to validate the population-genetic algorithm.

Restriction of the patient cohort to the patients of European origin

Most of the patients (9038/12048, 75%) were of European origin. The control population was formed by the gnomAD database, where a total of 138 632 genotype data are available. Because the Caucasians were the highest proportion in both the gnomAD database and our patient population, we examined the variants carried by non-Finnish persons in both the patient and control populations in Europe: variants of 63 369 European non-Finnish people included in gnomAD were compared with variants of 8805 European non-Finnish patients in the patient population. We included one patient per family for further calculations (n = 233 siblings were excluded). Thus, 2498 variants of 8805 unrelated patients were investigated.

Compensation of inbreeding

Inbreeding results in the overrepresentation of the homozygous patients and bias the allele frequencies in the patient populations with different degrees of inbreeding. To compensate for its effect, we calculated the coefficient of inbreeding for each variant according to the equation. This compensation resulted in the elimination of 287 variants that were all present only in the homozygous state in a single individual.

Validation of variant pathogenicity

To avoid the misinterpretation of benign HGMD variants as IP, we validated the pathogenicity of all HGMD variants on a population-genetic basis. We considered all variants to be pathogenic if they were significantly ($p < .05$) more frequent in the European, non- Finnish patient population than in the European, non-Finnish gnomAD population. Variants that were less frequent in the patient population were considered benign and excluded from further analyses. Along the same line, variants that were not significantly enriched in the

patient population were also excluded from further analyses. After the exclusion of the non-enriched and the non-significantly enriched variants, patients without biallelic variants were excluded from further analysis (228/6797 patients). We refer to the remaining 6569 patients of European origin as the patient population and the non-Finnish European gnomAD population as the control population.

Exclusion of disorders that are lethal in utero

Variants may be underrepresented in the patient population and thus mimic incomplete penetrance secondary to in utero lethality. To assess its possibility, we compared the proportion of the LOF and the 1476 missense pathogenic variants between the patient and the general population by Fisher's exact test and the Hardy–Weinberg equilibrium of missense and LOF variants within the patient population. The limit of significance was set at p = .00294 (= 0.05/17 comparisons) for both analyses. Disorders that suggested in utero lethality by both approaches were excluded from further analyses.

Penetrance calculation

We estimated the penetrance of 1936 variants, which were all significantly enriched in the patient population. We considered the LOF variants to be completely penetrant in the selected AR disorders with no fetal demise and no unaffected sibling reported with biallelic LOF mutations. To assess the penetrance of each variant, we compared its relative enrichment in the patient versus the general population to that of LOF variants as follows: : $\frac{AC_{bet}^{V} / AC_{gnomAD}^{V}}{AC_{bet}^{LOF} / AC_{gnomAD}^{LOF}}$, where we denote the allele count of a variant V by ACpt V and ACgnomAD V in the patient and the general population, respectively. Similarly, ACpt LOFand ACgnomAD LOF denote the cumulative allele counts of the LOF variants in the patient and the general population

Association tests

To identify interallelic interactions as the cause of IP, we investigated whether IP variants are associated with specific (type of) variants in trans.
The trans-associated variants were categorized according to their domain localization or consequence (missense/LOF), and were compared to variants trans-associated to all other variants by Fisher's exact test.

Validation of the population-genetic approach

To validate the specificity of the population-genetic approach, we compared the clinical (hypomorphic/null) and the biological characteristics (missense/LOF) of the IP and completely penetrant (CP) variants by Fisher's exact test. For this comparison, we considered the variants with a >70% penetrance to be CP.

Investigation of NPHS2 associations by FRET method

Interallelic interactions identified by population genetic method were further examined experimentally. To understand the relationship between oligomerization and pathogenicity, we compared the structure of wild type and R229Q oligomers, respectively the R229Q variants pathogenic (A297V, A284V, R291W, P341S, F344Lfs*4 associations), and benign (V290M, wild type, R229Q associations) oligomers.

Fluorescence spectroscopy

Podocin-coding plasmids (pcDNA 3.1 Zeo+) were provided by our collaboration partners, amplified in competent bacteria, purified with plasmid extraction kits, controlled via Sanger sequencing. Podocin variants were expressed in HEK293 cells, cultured at 37 °C in DMEM, high glucose with 10% FBS and 1% Penicillin-Streptomycin. Cells were transfected with Calcium-phosphate based method. Transfected cells were incubated for 48 h and lysed by 150mM NaCl, 20mM Tris, 1% Triton-X supplemented with 0.1% protease inhibitor. Lysis and upcoming procedures were performed on ice. Lysates were incubated with monoclonal anti-HA antibodies, subsequently with Protein G magnetic beads. Immunprecipitates were washed three times with lysis buffer. Podocin variants were eluted by competition with HA peptides. Concentration of the eluates was measured by spectrophotometry. Eluates were verified by Western blot analysis. Protein aliquots (0.4 nmol) were stained separately with 4 nmol of either Alexa Fluor 488 C5 Maleimide (donor dye) or Alexa Fluor 555 C2 Maleimide (acceptor dye) molecules and were incubated overnight at 4 °C. Differently stained podocin variants were subsequently mixed two by two for oligomerization, incubated for 2 h on RT and washed through PD SpinTrap G-25 filter column to discard the unbound HA peptides and fluorophores. Förster type Resonance Energy Transfer (FRET) was measured between two differently stained podocin variants in a final volume of 100 µl, containing each podocin with a concentration of 4 µM. The fluorescence lifetime of the donor dye (Alexa 488 C5) was measured, out of which the FRET efficiency was counted, which refers to the binding capacity between the homo-end heterooligomers. Experiments were repeated three times on samples obtained from at least two different expressions.

Statistical analyses

To determine the pathogenicity, intrauterine lethality, penetrance, association tests and for population genetic method validation we used Fisher's exact test.
The comparison of FRET efficiency was examined using a Mann-whitney U-probe. For multiple comparisons, Bonferroni correction was applied.

Establishing a clinical guideline for assessment of NPHS2 R229Q associations

Based on different methods (family studies, population genetics, biological and biophysical approaches) we have created a clinical guide which helps to assess the pathogenicity of associations with the *NPHS2* R229Q variant.

Datas were used from studies previously published by our research team.

4. Results

Identification of benign variants

As non-pathogenic HGMD variants can be falsely interpreted as IP, we first aimed to exclude them. These were expected to be less frequent in the patient than in the general population. Out of the 2032 variants, we found 25 (1.23%) such variants. We also excluded from further analyses the variants that were nonsignificantly enriched in the patient population (n = 111).

Exclusion of in utero lethal disorders

We tested the Hardy–Weinberg equilibrium for missense and LOF variants in each disorder after compensation for inbreeding, and found a disequilibrium in the *DHCR7* variant. To unravel the underlying phenomenon, we compared the ratio of the pathogenic missense and the LOF variants between the patient and the general populations, and found the *DHCR7* LOF alleles to be highly underrepresented in the patient population (7.7×, $p = 9.45 \times 10^{-76}$), suggesting that biallelic LOF alleles are lethal in utero in most of the cases, in accordance with former analyses associated.

We found the *NPHS2*, *ASL*, *PKHD1*, *CAPN3*, and *CFTR* missense alleles to be underrepresented in the patient population, suggesting the presence of IP missense variants in these genes.

Identification of IP variants

We considered a variant to be IP if it was enriched in the patient versus the general population, but its enrichment was significantly

($p < 7.22\times10^{-5}$) lower than that of the LOF alleles with a penetrance <30% in a disorder without prenatal demise. Out of the 1936 variants, only 85 was frequent enough in the general population to theoretically reach such a significant diminishment. We found 25 of them (29.4%) to be IP, including the three known IP variants (*CFTR* R117H, L997F and *NPHS2* R229Q and 22 novel ones. We identified IP variants in all genes with a reduced missense/LOF ratio in the European patient population.

Validation of the population-genetic approach

While the IP variants were typically hypomorphic, the CP variants were null in a majority ($p = 5.12 \times 10^{-10}$), indicating the good specificity of the approach. Along the same line, none of the IP variants, but 34.46% of the CP variants were a LOF variant ($p = 2.86 \times 10^{-5}$).

	hipo	**null**	**p (hipo vs. null)**	**nd**
ip (25)	16	2	$5{,}12\text{x}10^{-10}$	7
cp (1660)	177	756		727
	miss	**LOF**	**p (LOF vs. miss)**	**other**
ip (25)	**23**	**0**	$2{,}86\text{x}10^{-05}$	**2**
cp (1660)	924	572		164

Hunting for interallelic interactions

Incompletely penetrant variants were screened for interallelic interactions based on their unequal trans-associations to specific variants.
A single variant, *NPHS2* R229Q corresponded to the criteria of being subject to a dominant negative effect. Its overrepresented associations

in the patient population are indeed well-known pathogenic associations (A284V, A288T, A297V, E310K, E310V, Q328R, F344Lfs*4).
None of the variants was found to be the subject of complementation.

Structural examination of podocin oligomers

Comparing the wild type and R229Q homo-and heterooligomers, we found a significant higher FRET efficiency in case of R229Q associations ($p = 2{,}7x10^{-07}$).
Comparing the R229Q pathogenic (R229Q-A284V, R229Q-A297V, R229Q-R291W, R229Q-P341S, R229Q-F344Lfs*4) and benign associations (R229Q-vad, R229Q-R229Q, R229Q-V290M), we have seen that FRET efficiency was significantly higher in case of pathogenic associations than non-pathogenic associations ($p = 0.0029$).
On the basis of evidence reviewed above, both R229Q variant and variants with dominant negative effect increase FRET efficiency, which reflects the change in the structure of oligomers, confirming the relationship between pathogenicity and structural change.

A guide for clinical assessment of NPHS2 R229Q associations pathogenicity

Based on the presented observations, we propose to consider a mutation to be pathogenic with p.R229Q if it meets all of the following criteria:
1. The mutation fulfills the standard criteria of pathogenicity, that is, affects an evolutionary conserved amino acid.
2. The affected amino acid is located in the region of oligomerization (residues 270–351) and its substitution results in change of size, polarity, or hydrogen bonding capacity.
3. It does not disrupt the oligomerization

4. The expected frequency of individuals with [p.R229Q];[mut] in the general populations is lower than $1:10^6$.
5. It is enriched (even non-significantly) among patients with [p.R229Q];[mut] as compared with the patients with [mut];[mut], that is, a mutation of a novel p.R229Q association should not have been reported to be associated to another *NPHS2* mutation in more than a single case in populations with frequent p.R229Q.
6. The [p.R229Q];[mut] association segregates with the disease in the family.
7. The associated phenotype corresponds to the p.R229Q-associated nephropathy with no overt edema, FSGS on histology and low rate of progression (leading to ESRD between 10 and 50 years of age).

5. Conclusions

1. We developed a computerized algorithm which identifies frequent IP variants and interallelic interactions in large patient cohorts.

2. We found 25 variants, 29% of the frequent 85 variants, to be underrepresented in the patient population. Among them, only the *NPHS2* R229Q variant was subject to interallelic interactions.

3. We confirmed experimentally the role of abnormal oligomerization in the background of interallelic interactions of the *NPHS2* R229Q variant. Based on FRET efficiency, the R229Q podocin oligomers differ in structure from wild type oligomers, and pathogenic R229Q oligomers differ from benign oligomers.

4. We developed a clinical guideline for the assessment of unknown pathogenic associations of the *NPHS2* R229Q variant.

6. List of publication

Publications related to the thesis

Ágnes Mikó, Ambrus Kaposi, Karolina Schnabel, Dániel Seidl, Kálmán Tory
Identification of incompletely penetrant variants and interallelic interactions in autosomal recessive disorders by a population-genetic approach
HUMAN MUTATION 42: 11 pp. 1473-1487. (2021)
IF: 4,878*

Ágnes Mikó, Dóra K. Menyhárd, Ambrus Kaposi, Corinne Antignac, Kálmán Tory
The mutation-dependent pathogenicity of NPHS2 p.R229Q: A guide for clinical assessment
HUMAN MUTATION 39: 12 pp. 1854-1860. (2018)
IF: 4,453

Pál Stránera, Eszter Balogh, Gusztáv Schay, Christelle Arrondele, **Ágnes Mikó,** Gerda L'Aunéb, Alexandre Benmerah, András Perczel, Dóra K. Menyhárd, Corinne Antignace, Géraldine Mollete, Kálmán Tory
C-terminal oligomerization of podocin mediates interallelic interactions
BIOCHIMICA ET BIOPHYSICA ACTA-MOLECULAR BASIS OF ISEASE 1864: 7 pp. 2448-2457. (2018)
IF: 4,328

Publications not related to the thesis

Mikó Ágnes, Lóth Szendile, Müller Judit, Lotz Bence, Rossitto Patrizio, Szabolcs Andrea, Benyó Gábor, Jávorszky Eszter, Tory Kálmán, Dezsőfi Antal
Arthrogryposis–renalis diszfunkció–cholestasis szindróma [Arthrogryposis–renal dysfunction–cholestasis syndrome]
ORVOSI HETILAP 163: 2 pp. 74-78. (2022)
IF: 0,540**

Szeri Flora, **Miko Agnes**, Navasiolava Nastassia, Kaposi Ambrus, Verschuere Shana, Li Qiaoli, F Terry Sharon, Boraldi Federica, Uitto Jouni, van de Wetering Koen,Ludovic Martin, Daniela Quaglino, Olivier M Vanakker, Tory Kalman, Aranyi Tamas
The pathogenic p.R391G ABCC6 displays incomplete penetrance implying the necessity of an interacting partner for the development of pseudoxanthoma elasticum
Paper: **DOI: 10.1101/2020.11.26.20236489, 20 p. (2021)**
Csak repozitóriumban hozzáférhető közlemény