

Functional validation of genetic variants

Mohamed Z. Alimohamed

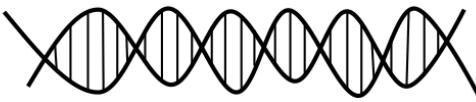
Introduction

- Currently, one of the main challenges in human molecular genetics is the interpretation of rare genetic variants of unknown clinical significance. A conclusive diagnosis is of importance for the patient to obtain certainty about the cause of the disease, for the clinician to be able to provide optimal care to the patient and to predict the disease course, and for the clinical geneticist for genetic counseling of the patient and family members. Conclusive evidence for pathogenicity of genetic variants is therefore crucial. This review gives an introduction to the problem of the interpretation of genetic variants of unknown clinical significance in view of the recent advances in genetic screening, and gives an overview of the possibilities for functional tests that can be performed to answer questions about the function of genes and the functional consequences of genetic variants (functional genomics)

What is functional genomics?

- Functional genomics is the study of how genes and intergenic regions of the genome contribute to different biological processes. A researcher in this field typically studies genes or regions on a “genome-wide” scale (i.e. all or multiple genes/regions at the same time), with the hope of narrowing them down to a list of candidate genes or regions to analyse in more detail.
- The goal of functional genomics is to determine how the individual components of a biological system work together to produce a particular phenotype. Functional genomics focuses on the dynamic expression of gene products in a specific context, for example, at a specific developmental stage or during a disease. In functional genomics, we try to use our current knowledge of gene function to develop a model linking genotype to phenotype.
- There are several specific functional genomics approaches depending on what we are focused on:
 - DNA level (genomics and epigenomics)
 - RNA level (transcriptomics)
 - Protein level (proteomics)
 - Metabolite level (metabolomics)

Genomics and epigenomics



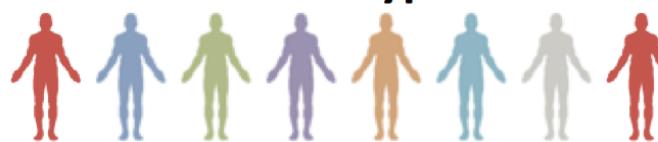
The study of the DNA sequence and associated heritable biochemical modifications

Transcriptomics



The study of the RNA molecules present in a sample

Phenotype

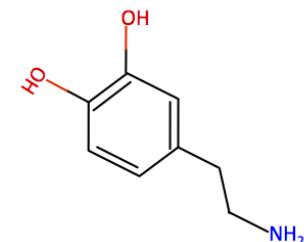


Proteomics



The study of the proteins present in a sample.

Metabolomics



The study of the metabolites present in a sample.

Common study types in functional genomics

- Functional genomics experiments measure changes in the DNA (genome and epigenome), RNA (transcriptome), or interactions between DNA/RNA and proteins that influence the phenotype of a sample. Common branches of functional genomics include:
- genotyping
- transcription profiling
- epigenetic profiling
- nucleic acid-protein interactions
- meta-analysis

Transcriptional profiling

- This is one of the most popular study types, also known as ‘expression profiling’. It involves the quantification of gene expression of many genes in cells or tissue samples at the transcription (RNA) level. Quantification can be done by collecting biological samples and extracting RNA (in most cases, total RNA) following a treatment or at fixed time-points in a time-series, thereby creating ‘snap-shots’ of expression patterns.
- For common reference genomes with well-annotated transcripts and genes (e.g. the human genome), a researcher can choose to focus on quantifying transcription of all or a subset of transcripts, genes, coding exons, non-coding RNA, and so forth.

Genotyping

- Genotyping studies are those which identify differences in the DNA sequence (i.e genotype) of a sample. The genomic DNA samples are often obtained from two contrasting groups of samples, e.g. case vs controls, with the aim of identifying differences in the genotype which may explain the difference in phenotype.

Genotyping studies can be designed to identify DNA sequence differences at three levels:

- Single nucleotide polymorphisms (SNPs, pronounced ‘snips’): SNP analysis focuses on differences in the DNA sequence at the single nucleotide level. Example: SNP 6.0 profiling of plasma DNA in breast cancer patients
- Copy number variations (CNVs): CNVs refer to an increase or decrease in the number of copies of a segment of DNA (e.g. a gene, or a locus-specific DNA repeat element). Each ‘copy’ can be as short as 50 bases or up to 100 kilobases. Example: A genome-wide copy number variant study of suicidal behavior
- Structural variations: they are an order of magnitude larger than CNVs and often cover megabases of DNA, and can be caused by chromosomal rearrangement events. Example: Transcription profiling by high throughput sequencing of different tissues to discover and characterize ‘presence-absence variation’ in the genome
- One common extension of genotyping studies in humans are genome-wide association studies (GWAS). Samples from cases (e.g. rheumatoid arthritis patients) and controls (e.g. healthy individuals) are genotyped across specific sites in the genome, followed by statistical analysis to find SNPs which are significantly more prevalent in one group (e.g. the disease cases). Such SNPs may then suggest an association between the SNPs and disease susceptibility.

Epigenetic profiling

- Epigenetics is the study of how biochemical modifications or physical interaction of DNA/chromatin affect gene regulation in a cell, where such modifications/interactions are not related to changes in the underlying DNA sequence.
- At the DNA level, methylation of CpG dinucleotides (often located near gene promoters) can be detected by first converting unmethylated cytosines into uracil using bisulfite, which allows methylated and unmethylated cytosines to be distinguished.
- At the chromatin level, modifications of the tails of histone proteins (e.g. methylation, acetylation) can be mapped by ‘immunoprecipitation’, where chromatin and proteins are chemically cross-linked reversibly. The genomic DNA associated with the modification/protein of interest is then ‘pulled-down’ (precipitated) with specific antibodies raised against the modification/protein. After precipitation, the cross-linking is reversed to release the genomic DNA for further analysis.

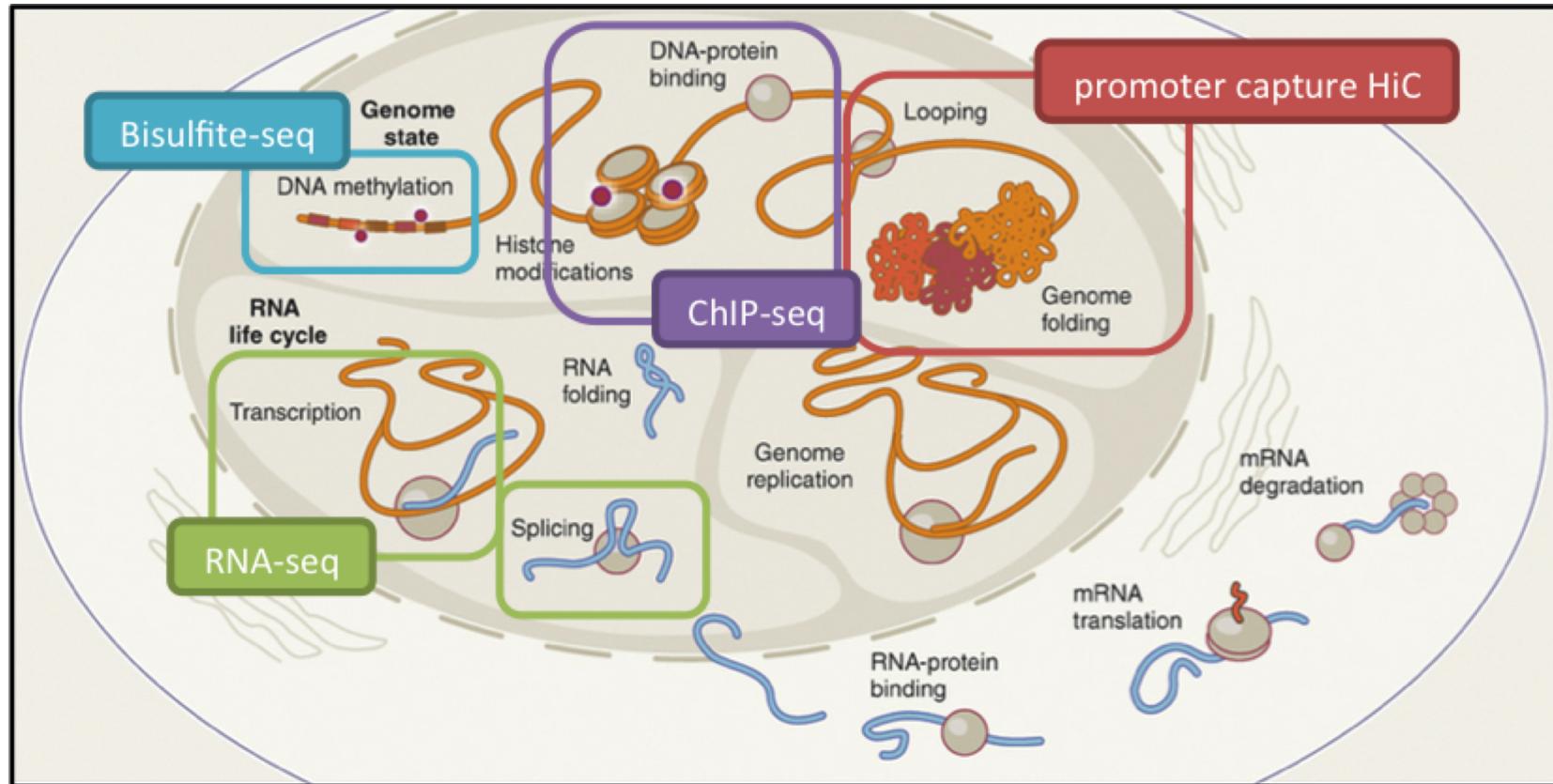
DNA/RNA-protein interactions

- Transcription factors, ribosomes and other DNA/RNA-binding proteins can bind to nucleic acid sequences and influence the transcription and translation of genes. The immunoprecipitation technique has also been applied to study protein binding sites on RNA.

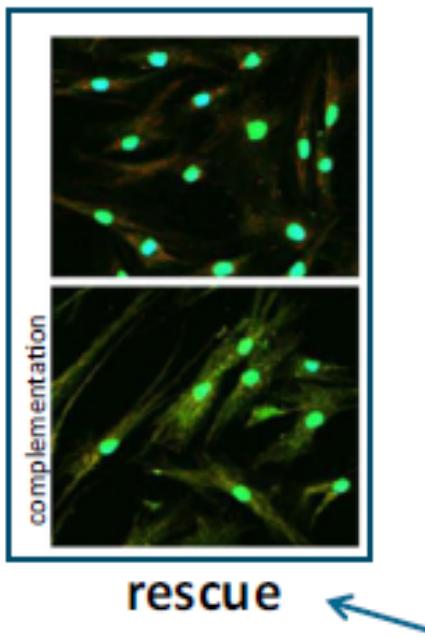
Meta-analysis

- Meta-analysis is a branch of functional genomics in which data from pre-existing experiments is combined to create statistically more powerful models of a biological process. This type of analysis has become popular as it allows the identification of subtle events that could not be detected in smaller studies.
- Functional genomics databases such as ArrayExpress and Expression Atlas play an important role in these studies as reliable, well annotated sources of functional genomics data.

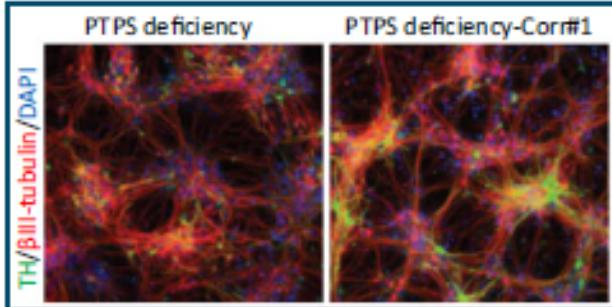
Molecules that can be analysed by functional genomics and the technologies used in the analysis



Examples of functional genomic approaches



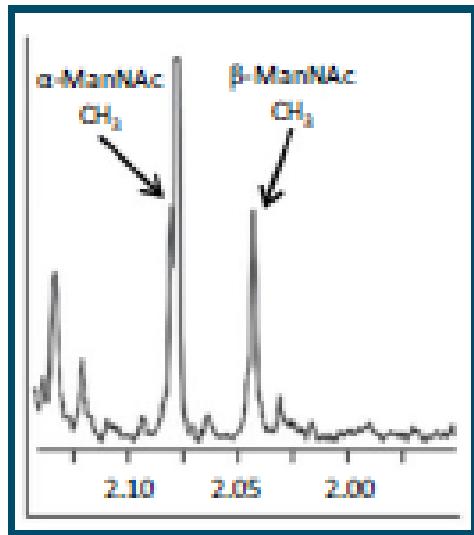
Rescue: introduction by lentiviral transduction of wild type LYRM7 cDNA in fibroblasts from a patient with a defect in LYRM7 results in normalization of mitochondrial Rieske Fe-S protein



CRISPR/Cas9



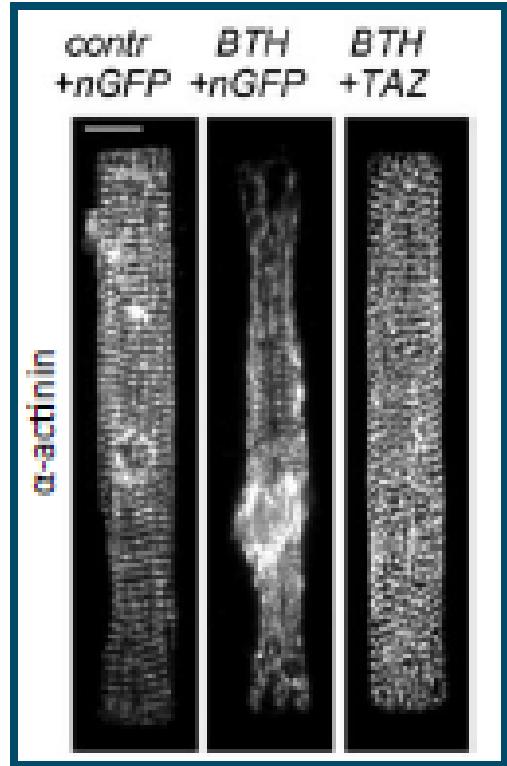
CRISPR/Cas9: Absence of thymidine hydroxylase (TH) staining in iPSC-derived dopaminergic nerve cells from a patient with a defect in PTPS, and normalization of TH expression after CRISPR/Cas9-mediated correction of the PTPS gene



biomarkers

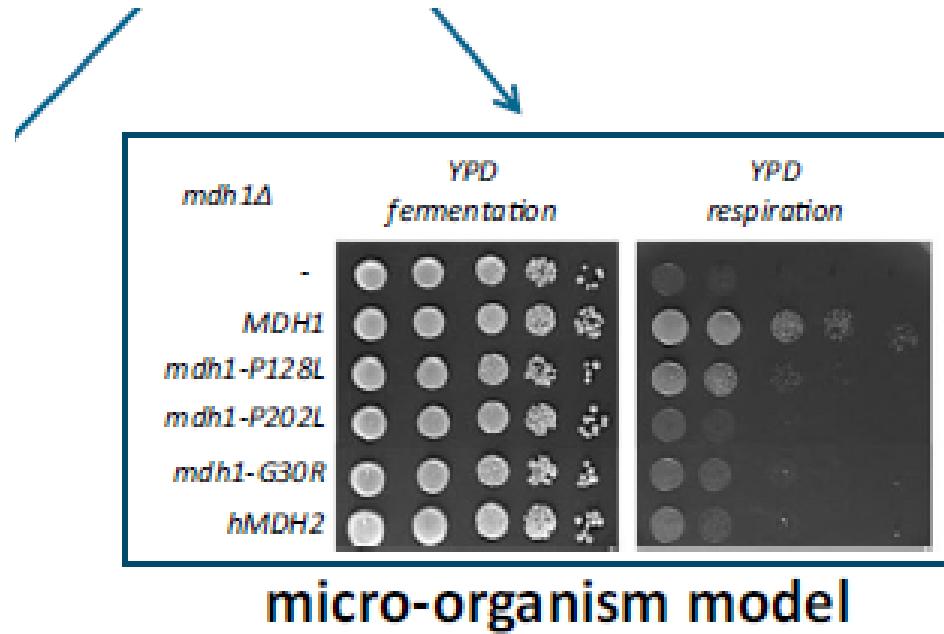


Biomarkers:
detail of a ^1H NMR spectrum of a CSF sample from a NANS patient,
showing the presence of alpha and beta forms of N-
acetylmannosamine

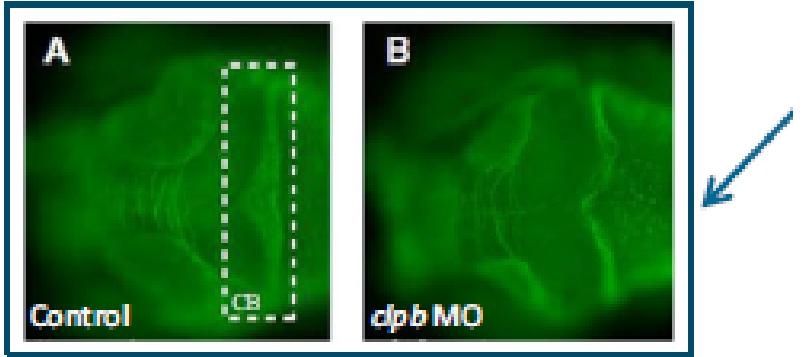


iPSC

iPSC: Abnormal sarcomere organization in iPSC-cardiomyocytes derived from fibroblasts from a Barth-syndrome patient (BTH) with mutations in the tafazzin gene (TAZ), and normalization of sarcomeres after transfection with TAZ-mRNA

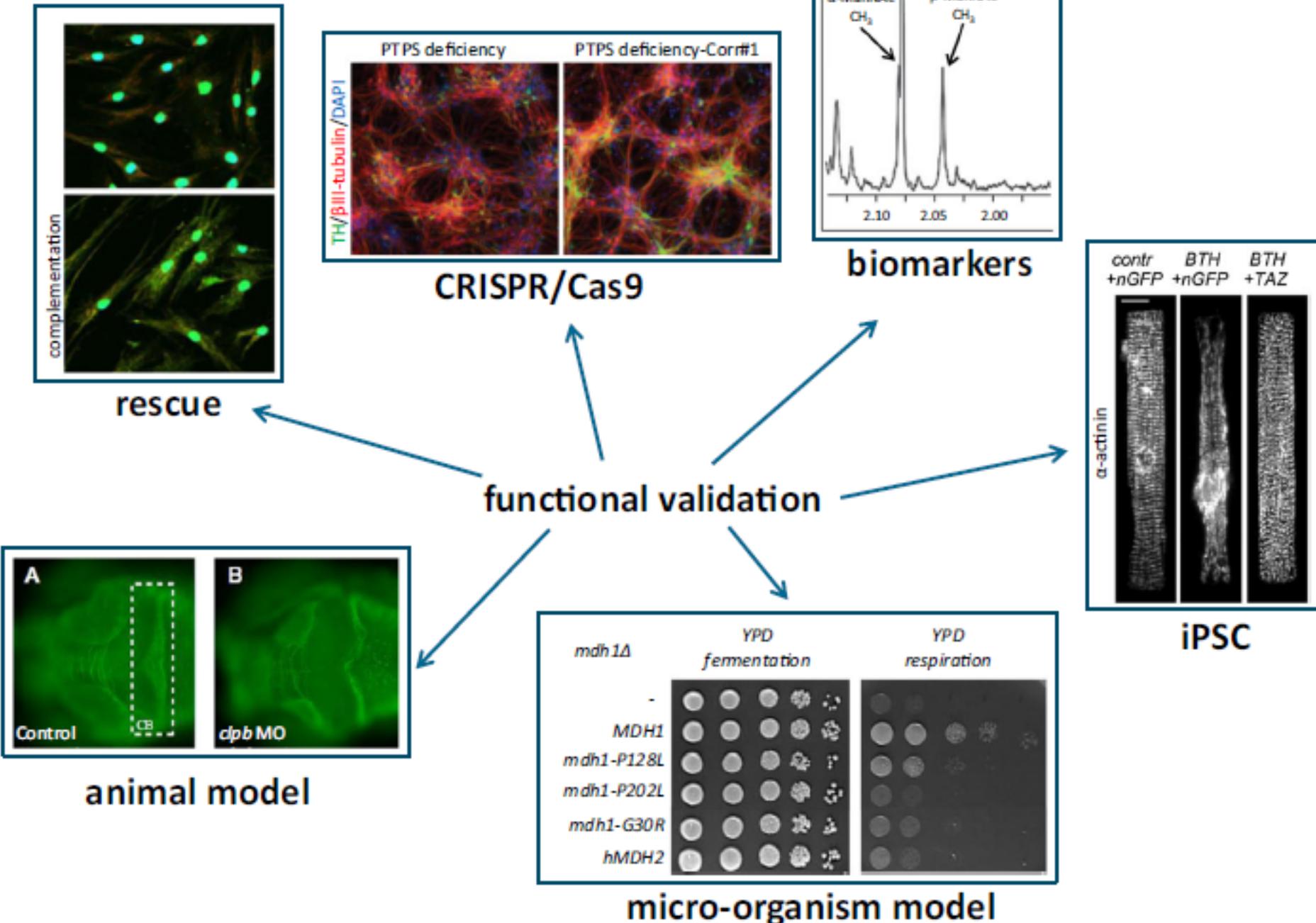


Micro-organism model: Restoration of growth on a nonfermentable carbon source (YPG) of a mitochondrial malate dehydrogenase-deficient yeast strain (*mdh1Δ*) by transfection with wild type MDH1, but not with various *mdh*-mutants



animal model

Animal model: A cerebellar defect in CLPB knock-down zebrafish embryos, as seen in Ac-tubulin stained embryos (the cerebellum is indicated by the rectangle in the control animal)



Summary of the main advantages and limitations of the most common technologies used for functional genome analysis

Technique	Advantages	Disadvantages	References
Variants detection methods			
GTG banding	Effortless analysis of the chromosome number and structure, including balanced rearrangements	Low sensitivity and resolution (5–10 Mb)	
FISH	Detection of minor structural cytogenetic abnormalities	Based on probes annealing to specific target	(11–13)
aCGH	High sensitivity and specificity	Inappropriate for the detection of balanced chromosomal rearrangements	
Sanger	High throughput, quality and reproducibility	Time consuming for large-scale projects	
NGS	Does not require a priori knowledge about genomic features Requires low amount of DNA/RNA as input	Expensive equipment. Complicated data analysis in the case of unspecified variants	(2, 19)

Epigenomics

Bisulfite conversion	Resolution at the DNA level. Effective method providing information about cytosine methylation	Impossible to distinguish methylated and hemimethylated cytosine (26–28)
MDRE	Easy to use Availability and assortment of endonucleases	DNA methylation assay is circumscribed by the use of a particular enzyme (28)
ChIP (including ChIP-chip, ChIP-seq)	Fast and well-studied. Compatible with array- or sequencing-based analysis, i. e., it is possible to perform genome-wide analysis	Relies on antibody specificity Microarray assay relies on particular probes (21)

Transcriptomics

Northern Blot	Quantitative and inexpensive method. No specialized equipment is needed There is a possibility of accurate display of the size and amounts of small RNA	Radioactive probes Lower sensitivity and lower throughput	(34)
SAGE	Direct and quantitative method. <i>A priori</i> knowledge about the gene sequences is not required. SAGE library requires a small amount of RNA as input. Simple data analysis	Low-throughput	(35)
qPCR	Fast, accurate, sensitive and highly reproducible method for mRNA quantification. Ability to detect the amount of mRNA in real time	The risk of bias	(36–38)
cDNA microarray	Well-studied, high-throughput and quantitative method Based on fluorescence (no need of radioactive probes)	Complicated data analysis	(39, 40)

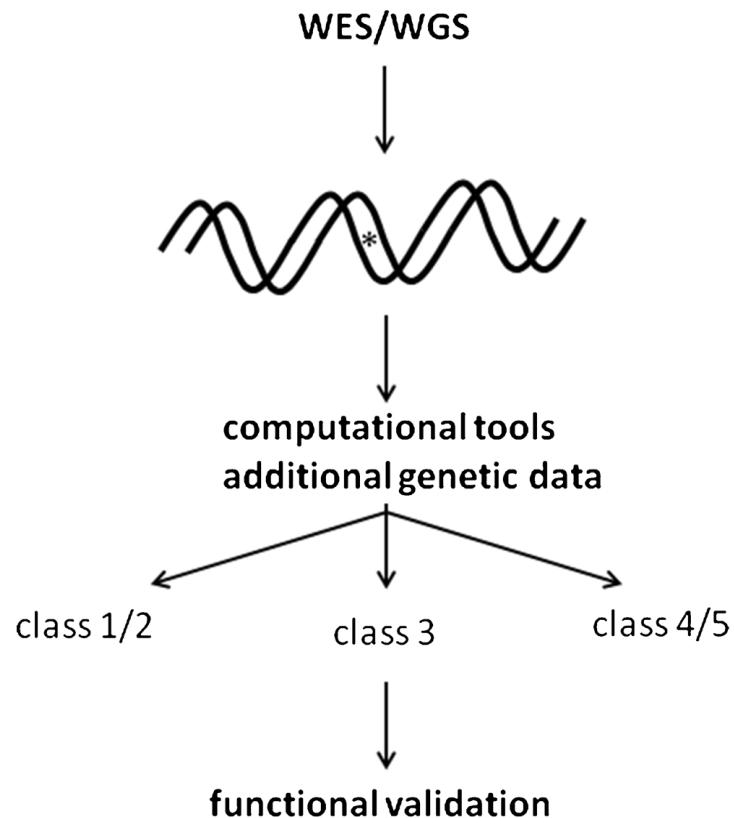
Transcriptomics

		(21, 42)
RNA-Seq	<p>Direct, quantitative and high throughput method. Does not require a priori knowledge about the genomic features.</p> <p>Appropriate for gene, transcripts (including alternative gene spliced transcripts) or allele-specific expression identification</p>	<p>High sequence similarity between alternative spliced isoforms</p>
Transgenesis of reporter gene	<p>“Gold standard” and accurate method for functional analysis of regulatory elements. Gene expression is easily detectable by fluorescence</p>	<p>Regulatory elements are widely dispersed through the genome that may cause some difficulties in detection</p>

Proteomics and interactomics			
ELISA	High sensitivity and specificity	Relies on antibody specificity	(47)
2-DE	Efficiently separates proteins by two properties	Poor separation of highly hydrophobic proteins. Inability to analyze very large or very small proteins	(48, 49)
MS	High-throughput method that rightly identifies and quantifies proteins	The sample should be high-quality and homogenous Sometimes the dissociation efficiency of complex protein is lower	(50, 52)
Y2H	The two-hybrid technique is relatively simple. Appropriate for the first step in identifying interacting protein partners	The rate of false positive results is relatively high. The need of confirmatory test. Impossible interaction between two proteins at a time	(7)

Model systems			
Chemical mutagenesis (in animal model)	Mutation can be induced artificially and mutant phenotype can be recognized easily Genes can be cloned using standard procedures	Phenotype not always reflects human beings	(61–63)
CRISPR-Cas9	The possibility to engineer the protein and RNA components of bacterial CRISPR system in order to recognize and cut DNA at desired locus	Work requires highly sterile conditions	(66)

From genetic test to functional validation

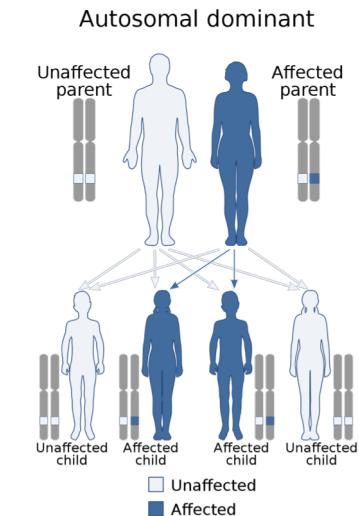
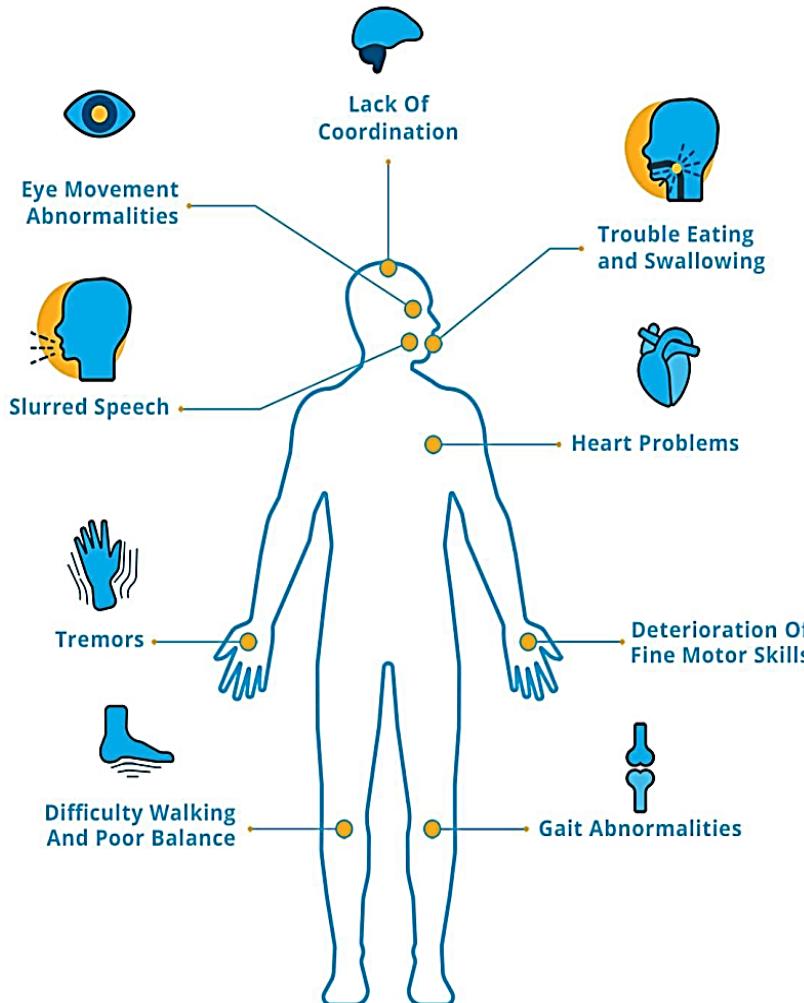


<u>model system</u>	<u>type of assay</u>
e.g.:	e.g.:
-patient cells	-genetic rescue
-blood	-overexpression
-model cell	-biomarker
line/organism	
-iPS cells	

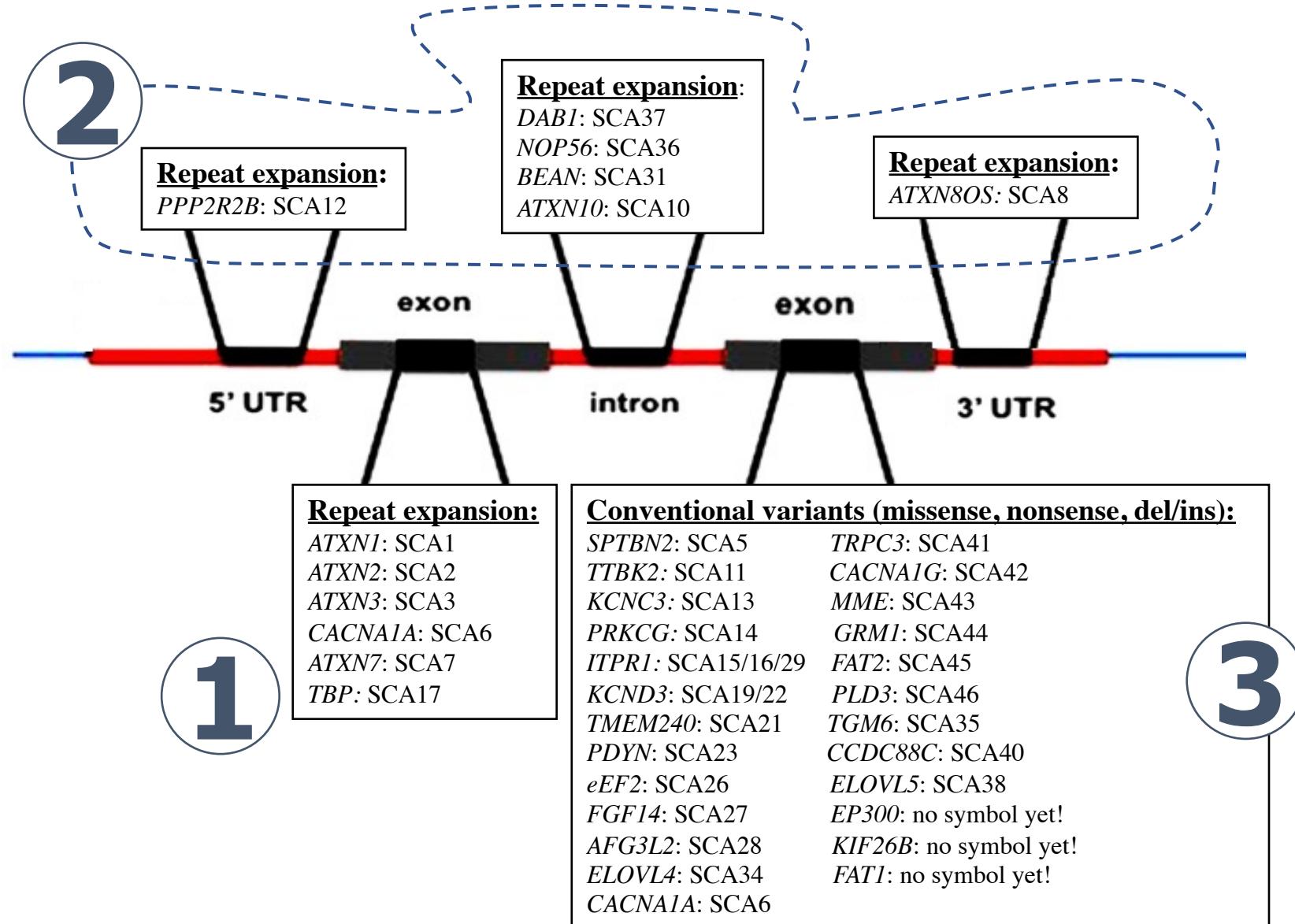
Clinical application example

Spinocerebellar ataxias (SCAs)

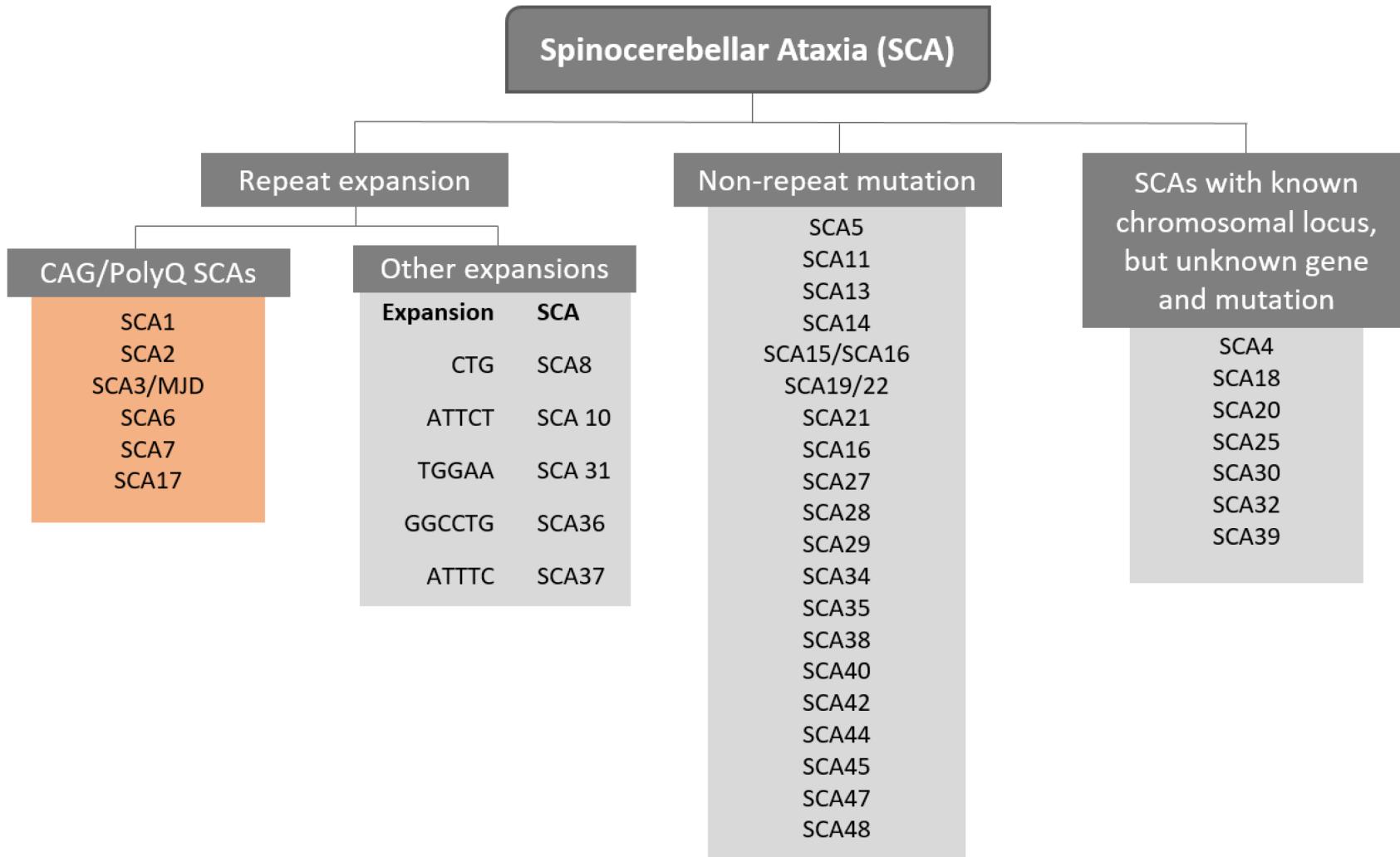
- Autosomal dominant neurodegenerative disorders.
- SCAs are a clinically and genetically heterogeneous group of rare disorders affecting approximately 1–5:100,000 persons worldwide
- To date 47 SCA types have been characterized for which 38 causal genes have been identified
- The most common is SCA3/MJD(Machado-Joseph disease) followed by SCA2 and SCA6, respectively



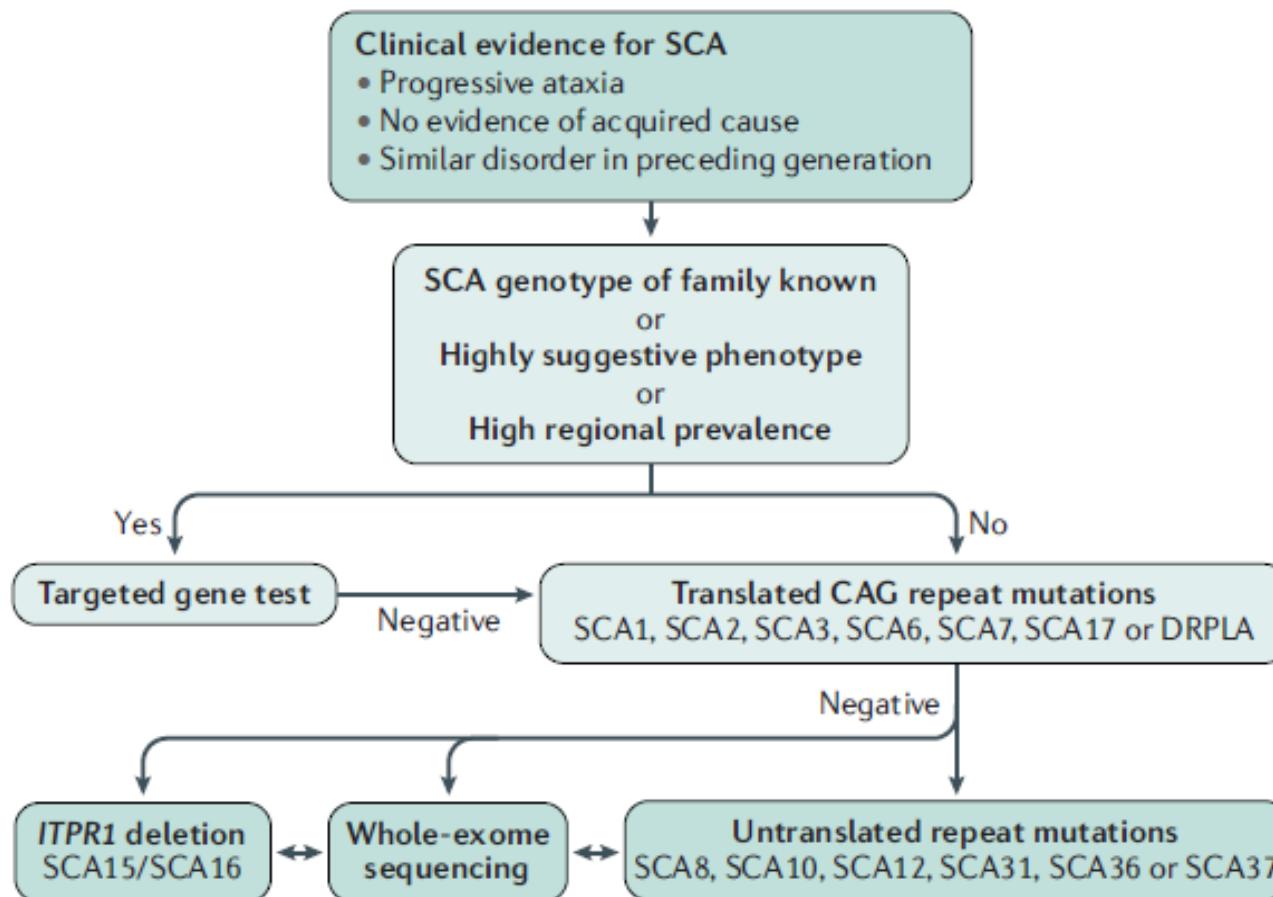
History of SCA genes and the mutations



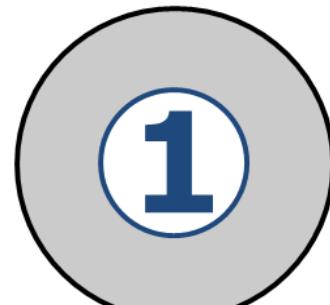
Representation of genetic classification of spinocerebellar ataxias



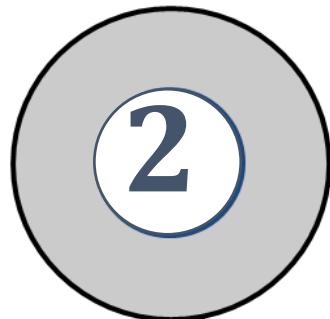
Flowchart of molecular genetic diagnosis of SCA



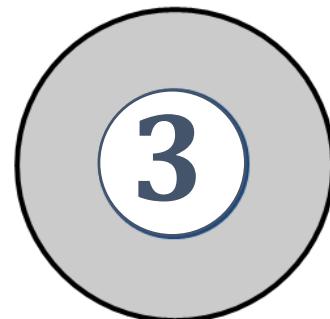
An undiagnosed SCA patient might carry....



A conventional mutation in non-coding regions
of **known SCA genes**



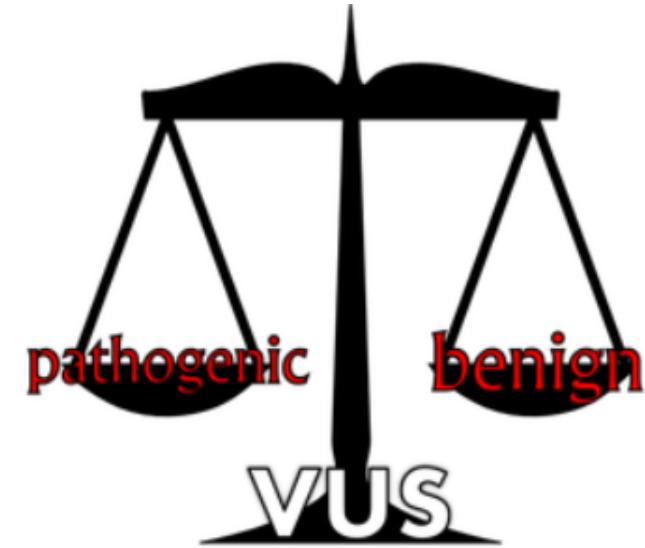
A **Variant of Unknown Significance (VUS)** in
known SCA genes which does have a significance!



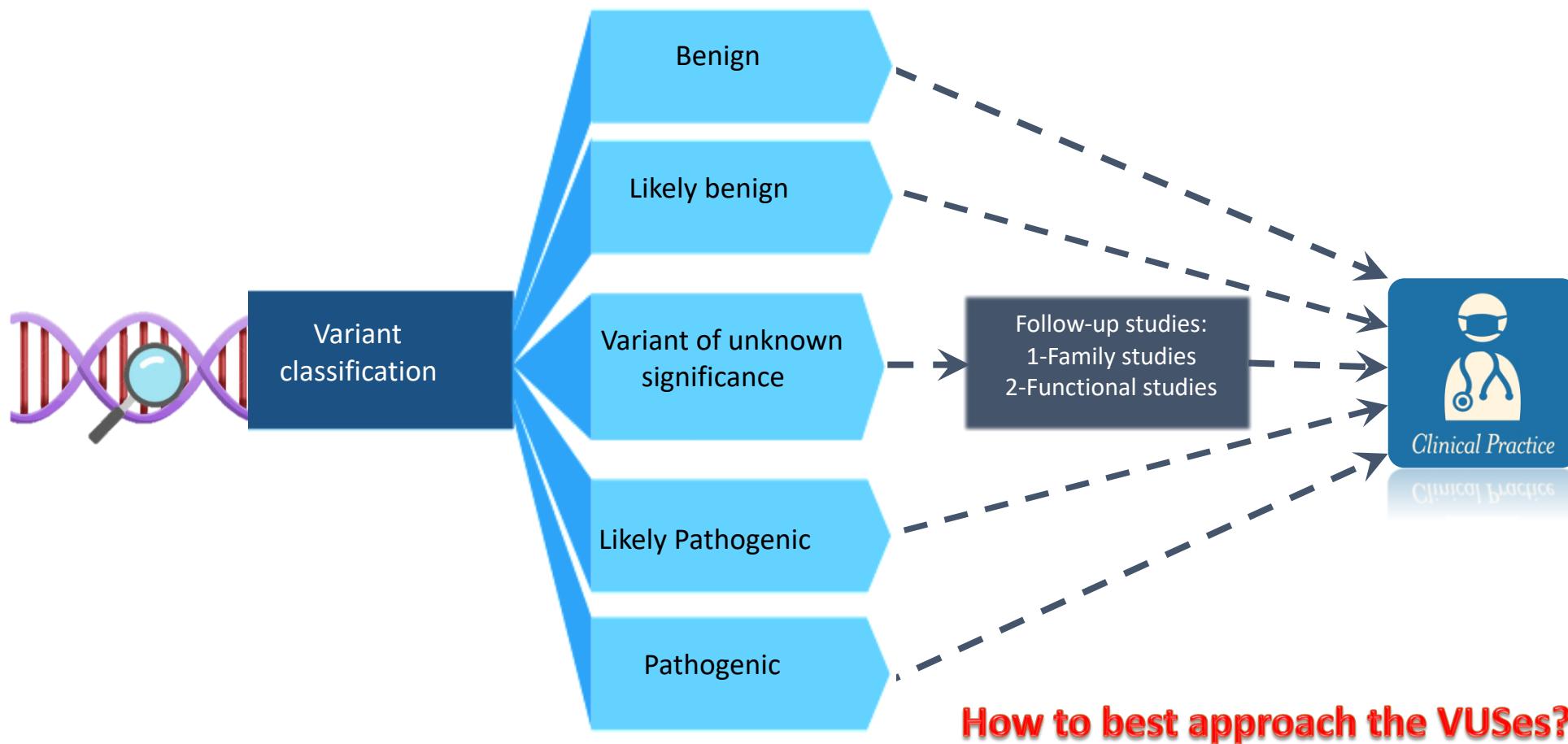
A mutation in **unknown SCA genes**

The problem

- Studies in ataxia patients have already shown that using Next Generation Sequencing (NGS)-based approaches for conventional variant detection can lead to a notable increase in the diagnostic rate
- Although the advent of NGS has enabled parallel sequencing of known SCA genes and the detection of conventional variants, it has brought with it the challenge of interpreting large numbers of variants with unknown significance (VUSes).



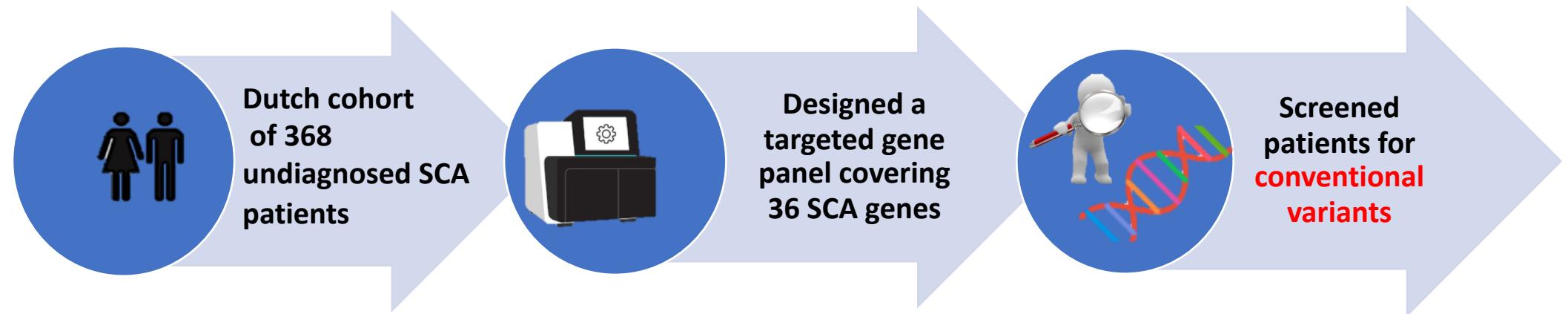
The challenge in clinical practice: Dealing with VUSes



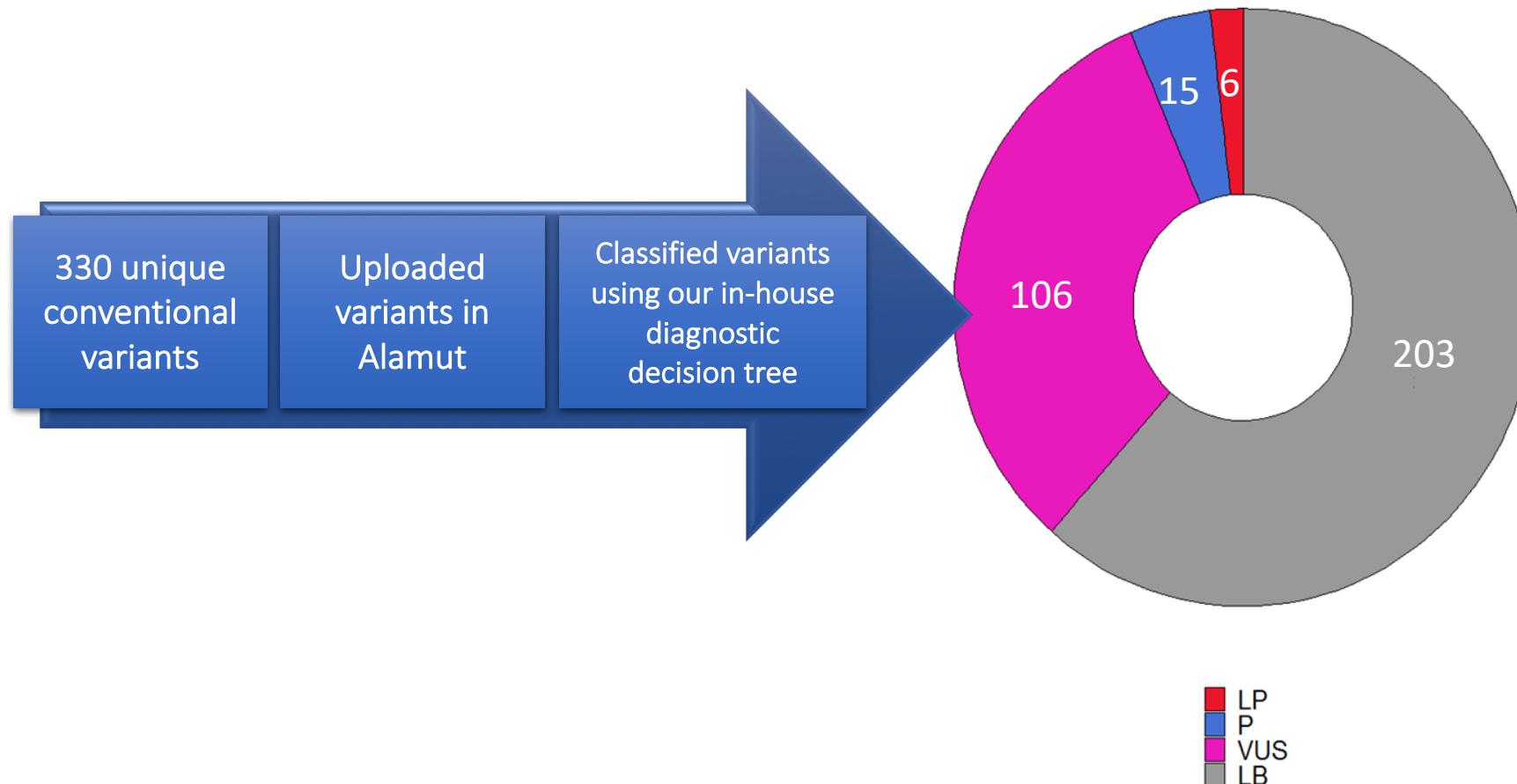
Our approach

- In this study, we addressed the challenge of prioritizing VUSes for follow-up in the context of SCA genetic diagnostics.
- We present a procedure for prioritization of VUSes for follow-up studies that include:
 - segregation analyses
 - protein modeling
 - functional testing and cellular testing

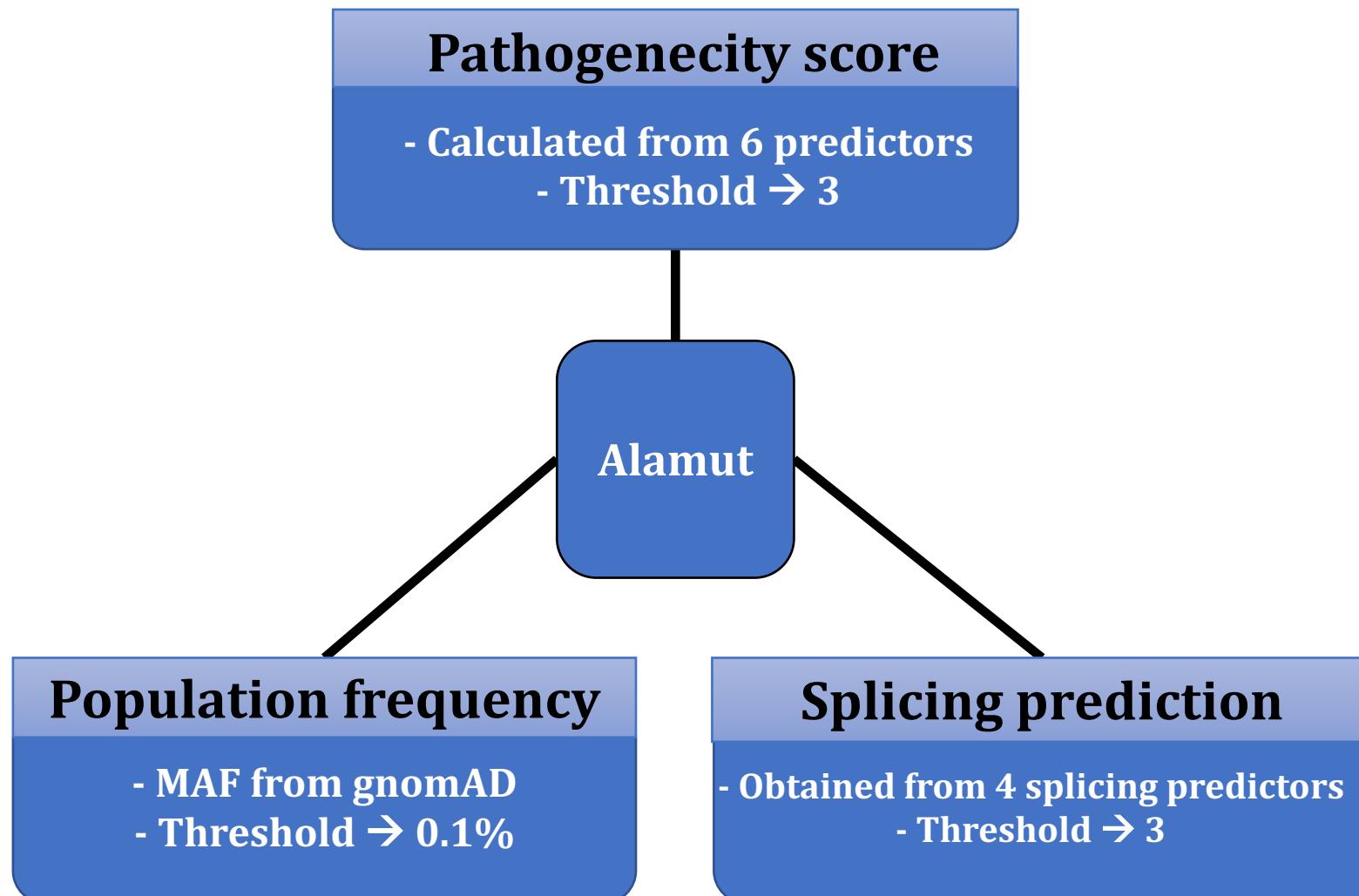
Dealing with VUSes in the context of SCA



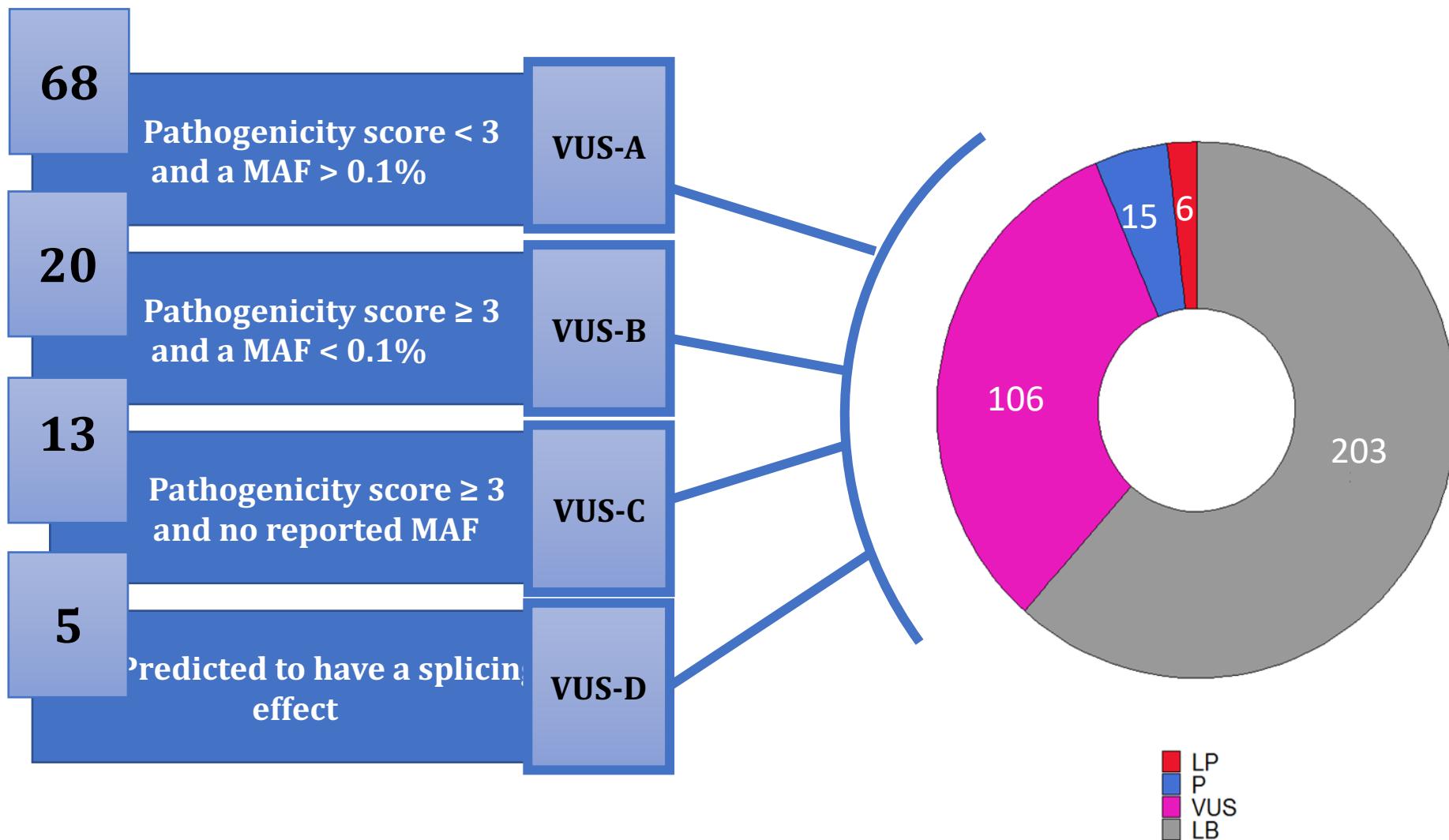
Variant Classification



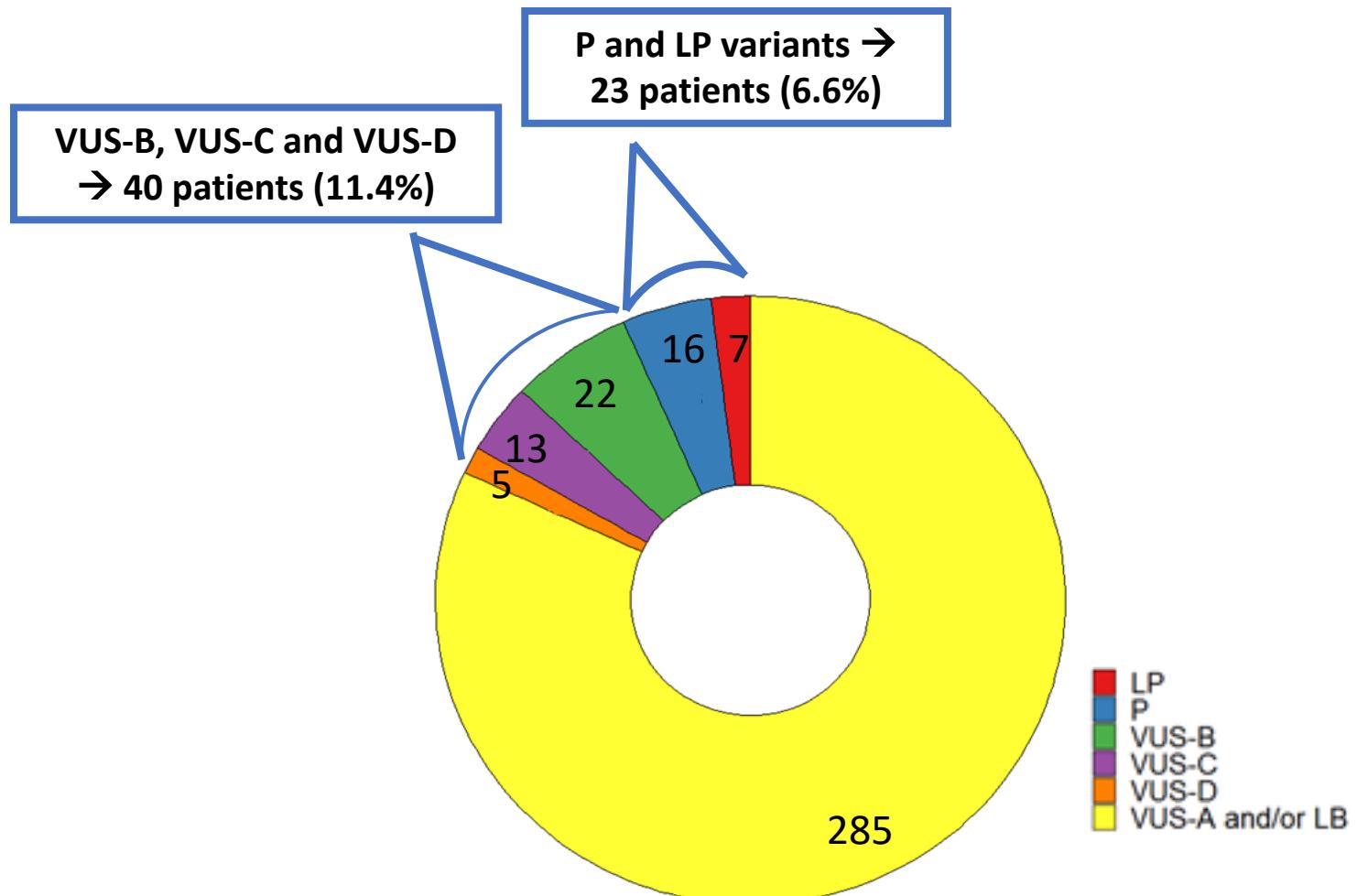
Parameters used for VUS sub-classification



VUS-subclasses



Distribution of VUS sub-classes, P and LP variants in the 348 patients



VUS sub-classification helps prioritize VUSes for follow-up studies

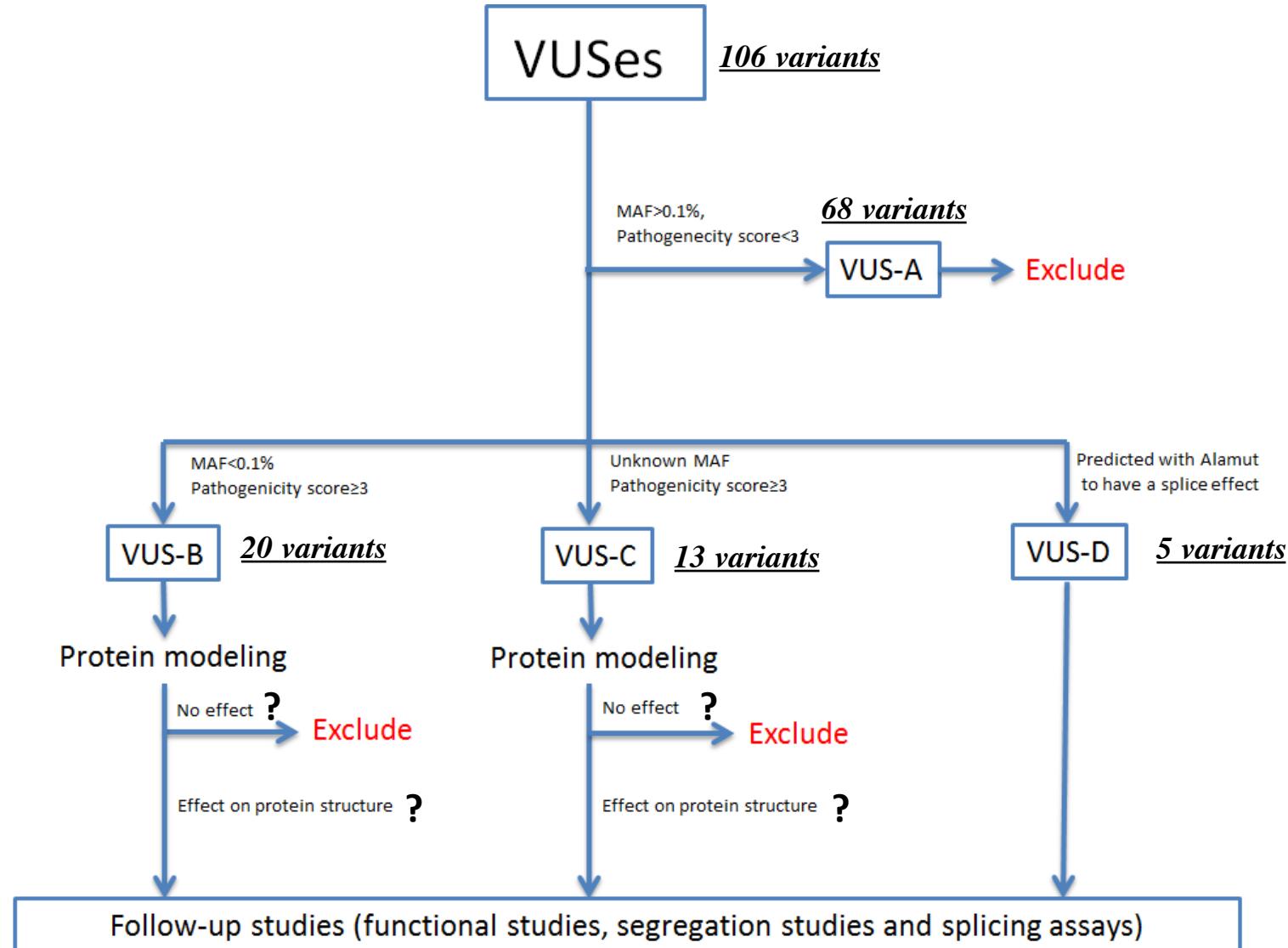
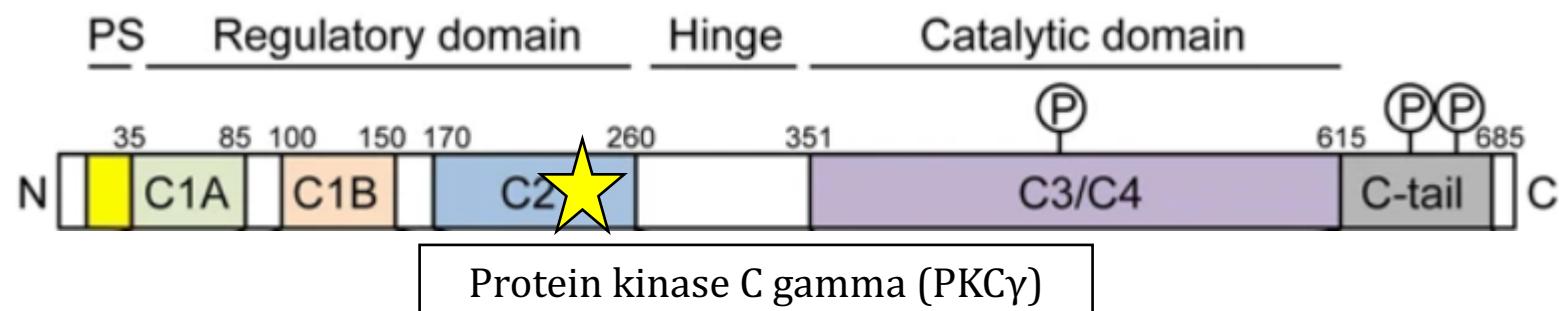
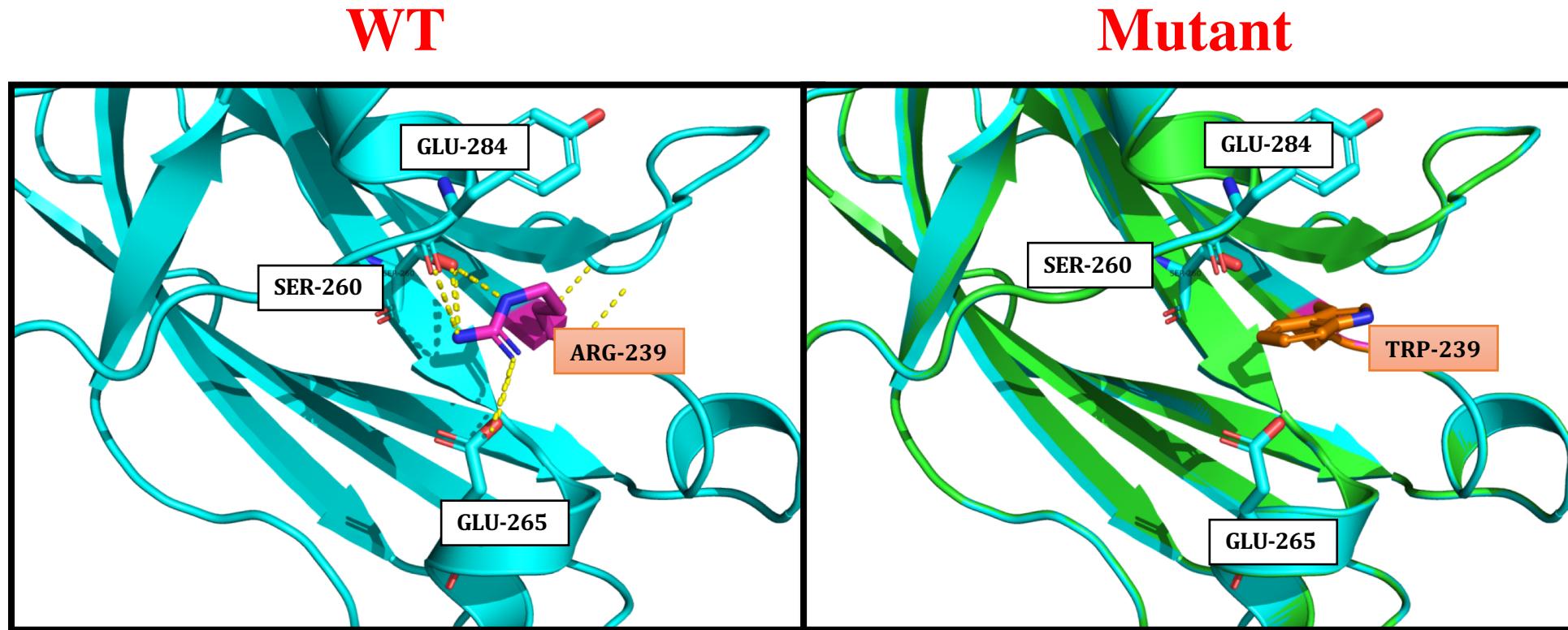


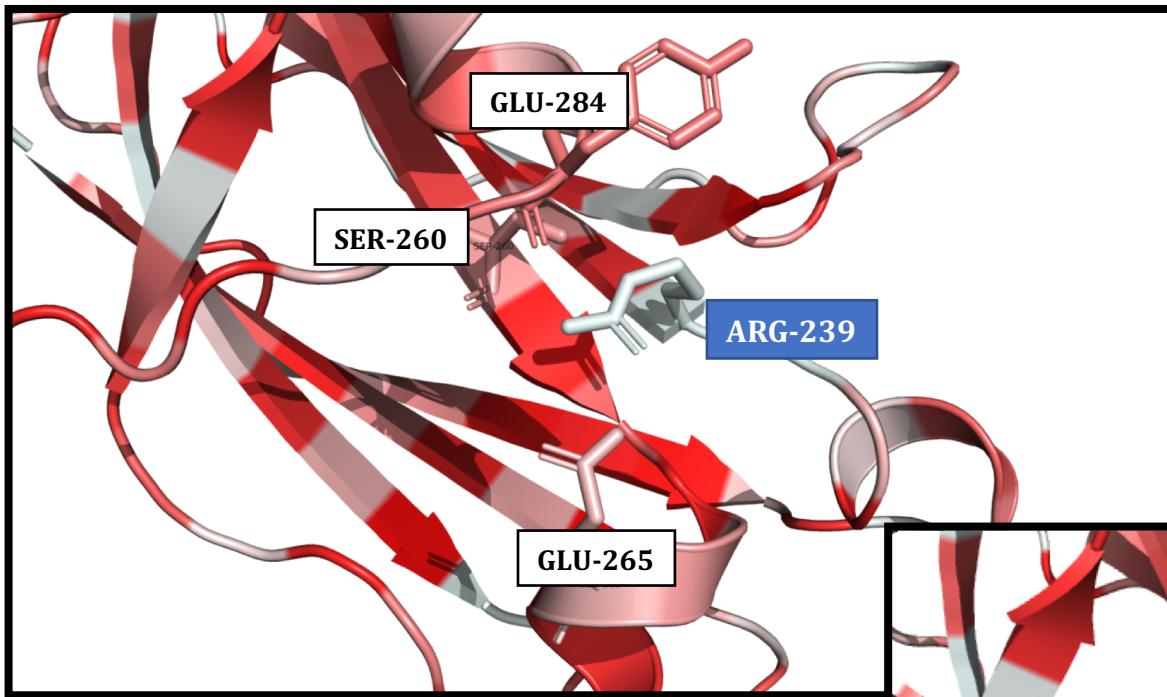
TABLE 1 | Proteins and the associated templates used for protein modeling.

Gene	Protein	UNIPROT code	PDB access code	PDB used for homology model	Sequence alignment (%)
<i>ARG3L2</i>	Afg3l2	Q9Y4W6	6NYY		
<i>CACNA1A</i>	Cav2.1	O00555		5GJW	49
<i>DAB1</i>	Dab1	O75553	1NTV		
<i>PRKCG</i>	PKCγ	P05129	2UZP		
<i>SPTBN2</i>	Sptbn2	O15020		1SJU	47
<i>TGM6</i>	Tgm6	O86932		3S3S	42
<i>PDYN</i>	Pdyn	P01213	2N2F		

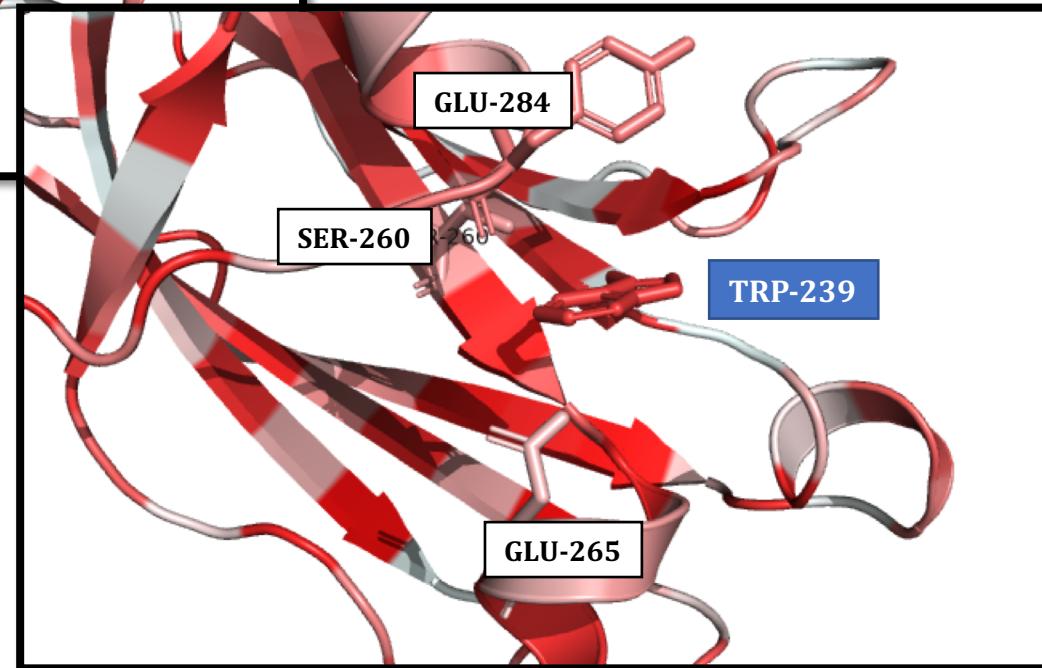
Protein modeling for *PRKCG*: c.715C>T:
p.R239W (VUS-C)



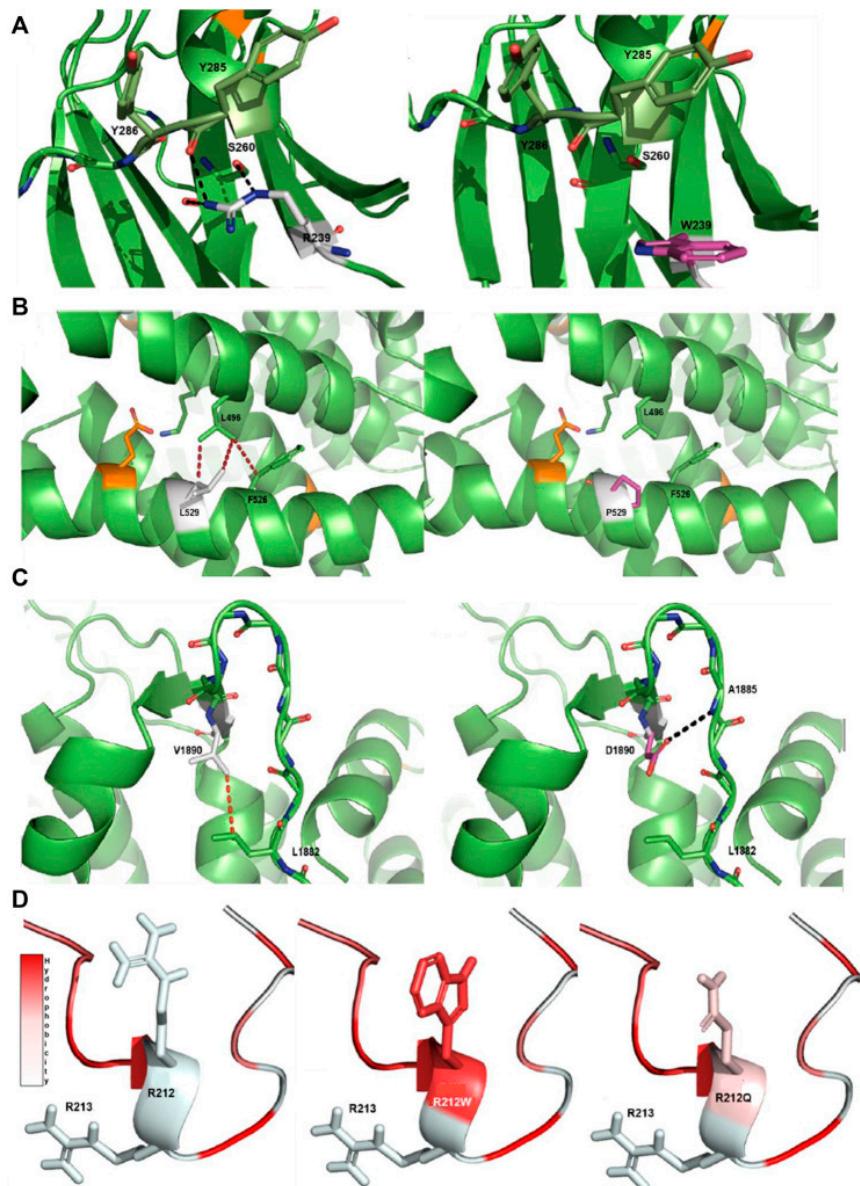
Changes in hydrophobicity



Red → hydrophobic regions
White → hydrophilic regions

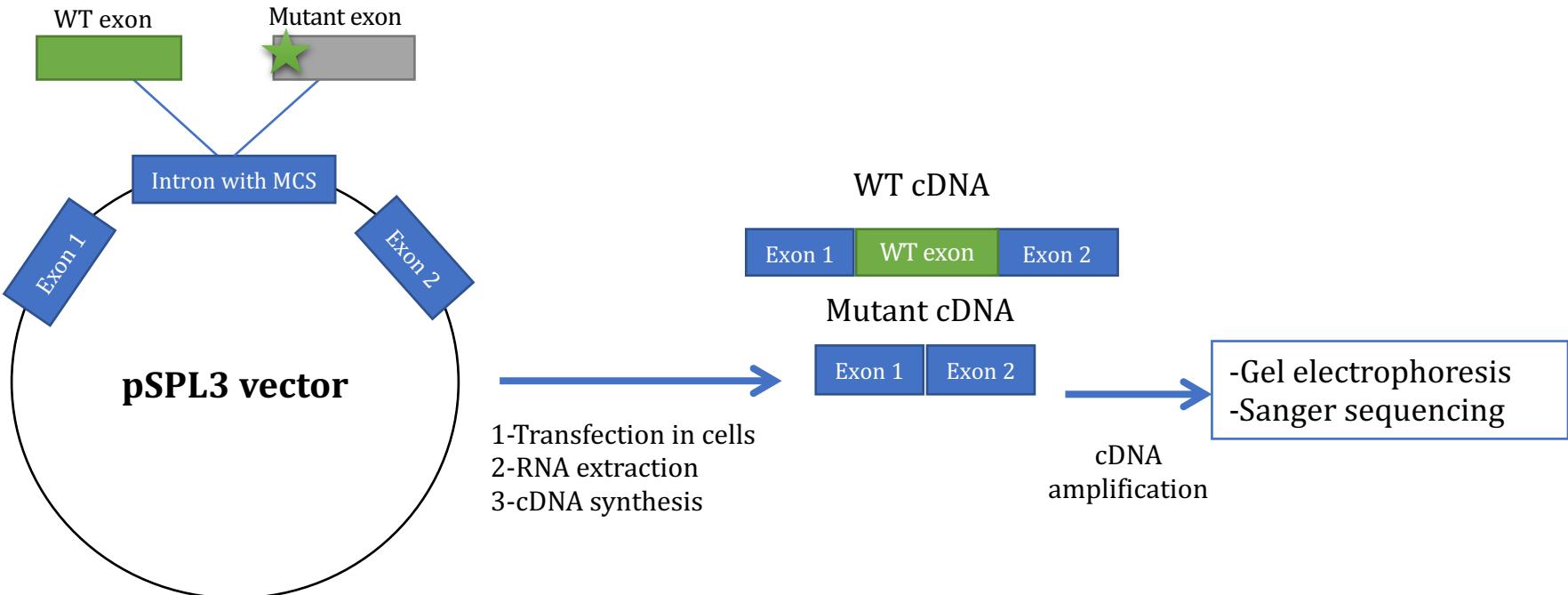


Protein modeling of variants with a notable effect on protein structure



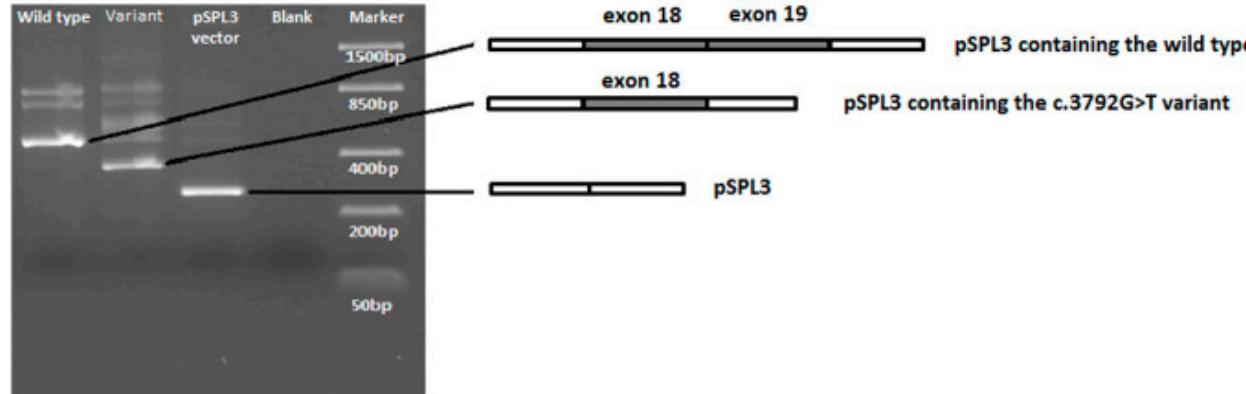
(A) R239W VUS-high in PKC γ protein. The native R239 interacts with S260 and Y285 through a hydrogen bonding network. The aromatic side chain of the mutant W is unable to establish an interaction network in the same manner as the native arginine. (B) L529P VUS-high in Cav2.1 protein. The native L529 engages in a network of hydrophobic interactions with L496 and F526. The introduction of P in the middle of the helix introduces a steric clash with these hydrophobic residues and is detrimental for helix stability (Visiers et al., 2000). (C) V1890D VUS-high in Cav2.1 protein. The native V1890 interacts with L1882 through a hydrophobic interaction. The mutated D changes the conformation of the backbone of the protein, disturbing the β -sheet structure and forming a new hydrogen bond with the main chain of A1885. Native side chains are shown as white, mutated side chains as pink, known pathogenic mutations as orange, hydrophobic interactions as red dashes and hydrogen bonds as black dashes. (D) R212Q VUS-semi high in pdyn protein. The R212Q mutation has similar hydrophobic properties to the known R212W, which is known to decrease the cleavage efficiency of dynorphin, and likely functions in a similar manner. Arginine (R), tryptophan (W), serine (S), tyrosine (Y), valine (V), aspartic acid (D), leucine (L), proline (P), glutamine (Q) and phenylalanine (F).

Confirming the predicted effect of VUS-D with the splicing assay



DNA gel electrophoresis of PCR-amplified cDNA fragments generated from wild type and variant sequences

A

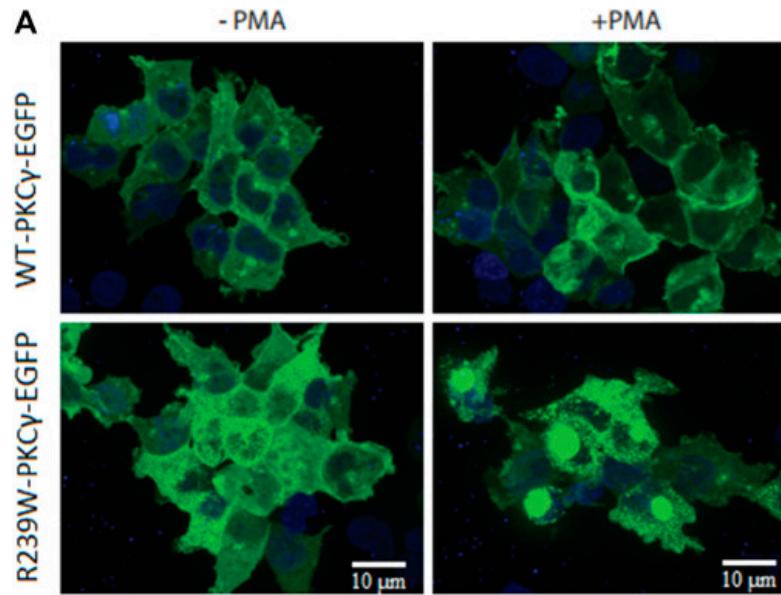


B



(A) For the CACNA1G c.3792G>T variant, the variant fragment is shorter and lacks exon 19 compared to the wild type fragment. (B) For NOP56, both the wild type and the c.909G>A variant sequences produce a fragment carrying exon 6 and 30bp of intron 6. However, the c.909G>A variant PCR product lacks the expected cDNA fragment carrying exons 6, 7 and 8.

Cellular studies of R239W PKC γ and V391M TGM6.



(A) While PMA-induced translocation was observed for wild type PKC γ -EGFP, very little translocation was observed for R239W-PKC γ -EGFP. (B) Immunocytochemistry did not show any difference in protein localization between wild type TGM6 and V391MTGM6.

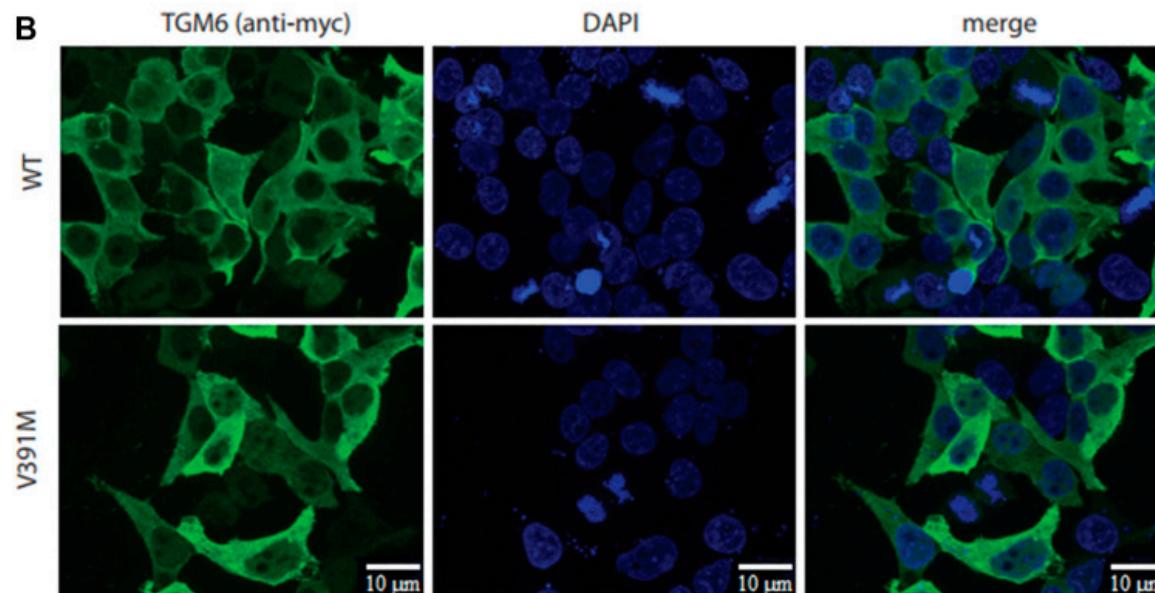


TABLE 4 | Reclassification of VUSes based on follow-up studies.

Patient ID	Gene	cDNA: Protein position	Subclass	Follow-up studies	Follow-up outcome	Final classification
38	ARG3L2	c.2143C>T: p.L715F	VUS-high	PM	PM predicted LB	VUS
28	CACNA1A	c.5669T>A: p.V1890D	VUS-high	PM	PM predicted LP	VUS
30	CACNA1A	c.2357G>C: p.R786P	VUS-high	SS	SS indicates LP	LP
31	CACNA1A	c.1586T>C: p.L529P	VUS-high	PM	PM predicted LP	VUS
32	CACNA1A	c.5157T>A	VUS-splice	FT	FT indicates LB	VUS
43	CACNA1G	c.3792G>T	VUS-splice	FT	FT indicates LP	LP
40	DAB1	c.209G>A: p.G70D	VUS-semi high	PM	PM predicted LB	VUS
27	FAT1	c.8991G>A	VUS-splice	FT	FT indicates LB	VUS
39	NOP56	c.909G>A	VUS-splice	FT	FT indicates LP	LP
4	PDYN	c.635G>A: p.R212Q	VUS-semi high	PM	PM predicted LP, segregates with known LP in CACNA1A	VUS
37	PRKOG	c.715C>T: p.R239W	VUS-high	PM, SS, FT	PM predicted LP, SS and FT indicate LP	LP
26	SPTBN2	c.1522A>C: p.N508H	VUS-semi high	PM	PM predicted LB	VUS
9	TGM6	c.1171G>A: p.V391M	VUS-semi high	PM, FT	PM predicted LB, FT indicates LB, segregates with known P variant in KCNC3	VUS

LB, likely benign; LP, likely pathogenic; PM, protein modeling; SS, segregation studies; FT, functional test.

Conclusions

- Our study shows that reclassification of VUSes is difficult due to the limited opportunities to carry out variant segregation studies and the limited weight and availability of protein modeling and functional tests
- We anticipate that this situation will not change on the near future because the AMCG classification guidelines are very stringent, diagnostic labs lack protein modeling experts and functional tests are not easily available
- We advise focusing first on the most clinically relevant genes, gaining expertise on the structure of these proteins, bringing trained protein modeling experts into diagnostics and making functional testing part of routine
- To improve this yield further, VUSes and available clinical, functional and protein modeling data should be shared more widely between labs in order to build sufficient knowledge to better weigh the value of those outcomes for classification and thereby give them a more prominent position in classification guidelines

Variant interpretation using literature based attributes of ACMG/AMP guidelines

Functional methods-PS3 and BS3

- To assess the validity of a functional assay, one must consider how closely the functional assay reflects the biological environment.
- A variant can be validated using the in vitro or in vivo functional assays to determine whether it is implicated in the disorder or not. For the variant validation, we can either use the patients' sample, or generate the mutation in cell lines or model systems using the gene-editing tools, or incorporate the mutation in the plasmid DNA for further analysis. The further analysis includes quantification of the mRNA or protein expression, determining the biological and cellular activity, or determining the phenotypic characteristics in the model organisms. The functional assays are performed for both the wild type and the mutant alleles and if there is a significant difference between the wild type and mutant, then according to the ACMG guidelines, that variant should be assigned PS3, while if the assay doesn't show significant difference between wildtype and mutant, then it will be assigned BS3 based on the ACMG guidelines.