CentoLSD - Variant classification at CENTOGENE

PROCESS

Every genetic variant identified at CENTOGENE goes through a comprehensive review process, following closely the American College of Medical Genetics (ACMG; https://www.ncbi.nlm.nih.gov/pubmed/25741868) guidelines. This classification scheme is greatly influenced by CENTOGENE knowledge in the field of human rare disorders. CENTOGENE has optimized the variant classification, implementing a new standalone criterion, namely PVS2. This very strong pathogenic criterion is assigned to variants that are confirming a deleterious effect via in vivo measurements of biomarker levels. For example, a variant associated with min. 20 ng/ml glucosylsphingosine (Lyso-Gb1) is classified as pathogenic, based on the internal validated observations that is a highly specific and 100% sensitive indicator to diagnose Gaucher patients. All identified and annotated variants undergo a medical validation regarding their possible relevance to the provided clinical symptoms and / or suspected diagnosis of the patient.

At CENTOGENE variants are mainly classified using five- tiered scheme: pathogenic, likely pathogenic, variant of uncertain significance (VUS), likely benign and benign. The variant classification is based on the likelihood to predispose or cause the observed phenotype/ disease. To strengthen this decisive topic CENTOGENE has created its own internal database that includes detailed clinical information, frequency and geographic origin. The depth of available information in this unique cohort allows in many circumstances a pathogenic/likely pathogenic or benign/ likely benign classification exclusively based on our internal data.

VARIANT CLASSIFICATION WORKFLOW

The technical validated genetic variants are evaluated in the clinical context. This classification workflow is applicable to variants in all Mendelian genes, and only clear observations are assigned.

Pathogenic and likely pathogenic variants: rare genetic variants (MAF <1%; using multiple databases like gnomAD, ESP, 1000 Genome, in-house variant database) are categorized initially by analysis of the functional impact or predicted effect on the protein (truncating or null variants, missense or in-frame, etc.). The variants are annotated with pathogenic annotations as identified in Human Gene Mutation Database (HGMD), ClinVar, gene-specific databases and in-house database. The output from in silico protein, splicing predictions and evolutionary / conservation data are reviewed. The clinical review implies in –depth analysis of the variant for segregation, association with clinical picture of the patient and family history, co-occurrences with other pathogenic variant(s) in the same gene or other genes of interest.

Classification scheme for pathogenic and likely pathogenic variants			
PVS1-PVS2	PS1-PS4	PM1-PM6	PP1-PP5
FAST IDENTIFICATION CODES FOR ACMG RULES			
PVS1= Null/ Truncating **PVS2= Internal disease-specific biomarkers**	PS1 = Same amino acid PS2= Confirmed de novo PS3= Functional assays including internal biochemistry PS4= Affected>controls	PM1= Hot spot/ critical domain PM2= Absent in controls PM3= AR and in trans PM4= Protein lengths change PM5= Different amino acid PM6= Assumed de novo	PP1= Segregation PP2= Low rate benign missense PP3= In silico/ computational PP4= Phenotype/ family history PP5= Reputable source
REQUIRED CRITERIA AND COMBINATIONS FOR PATHOGENIC AND LIKELY PATHOGENIC VARIANTS			
Pathogenic variants			
1 PVS1	min 1		
1 PVS1		min 2	
1 PVS1		1	1
1 PVS1			min 2
	min 2		
	1	min 3	
	1	2	min 2
	1	1	min 4
1 PVS2			
Likely pathogenic variants			
1		1	
	1	min 1	
	1		min 2
		min 3	
		2	min 2
		1	min 4

Benign and likely benign variants: the genetic variants with MAF >5% (using multiple databases like gnomAD, ESP, 1000 Genome, in-house variant database) are automatically assigned for BA1. Additionally, benign annotations are assigned for variants detected in healthy controls, co-occurrences with a pathogenic variant in the same gene or in another gene that clearly explains the phenotype of interest, no impact on function observed in appropriate functional assay, no segregation, or other data supporting benign classification (predictions, weak conservation, external literature).

Classification scheme for benign and likely benign variants		
BA	BS1-BS4	BP1-BP7
FAST IDENTIFICATION CODES FOR ACMG RULES		
BA1= Allele frequency >5%	BS1= Allele frequency> disease frequency BS2= Observed in healthy BS3= Functional assay against (including internal biochemistry) BS4= No segregation	BP1= Missense in LOF BP2= Co-occurrences cis/ trans (AD/AR) same gene BP3= Repeat in- frame BP4= In silico/ computational benign BP5= diagnosis confirmed by other /variant/ gene) BP6= Reportable source benign BP7= Synonymous
REQUIRED CRITERIA AND COMBINATIONS FOR BENIGN AND LIKELY BENIGN VARIANTS		
Benign variants		
1		
	min 2	
Likely benign variants		
	1	1
		min 2

Uncertain variants: genetic variants with not enough evidences to score for pathogenic / likely pathogenic or criteria for benign and pathogenic are contradictory are assigned to uncertain class.

VARIANT CLASS DEFINITION

Pathogenic variants: A pathogenic variant is a well-established disease- causing DNA change in CENTOGENE's internal database and / or literature. The main evaluation criteria are represented by strong genotype-phenotype correlations, independent confirmatory observations, and supporting pathogenicity functional assays. Classification as pathogenic is additionally assigned to variants that are confirming a deleterious effect via in vivo measurements of enzymatic activity and / or biomarker levels.

Likely pathogenic variants: A likely pathogenic variant is considered the probable cause of the patient's phenotype, or the effect on the protein function is predicted to be likely deleterious. Classification as likely pathogenic is additionally assigned to loss of function (LOF) variants detected in the genes related to metabolic disorders with no in vivo measurements of enzymatic activity and / or biomarker.

Variants of uncertain significance (VUS): An uncertain variant is a genetic variant with unknown or questionable impact on a particular clinical phenotype. The variant is typically very rare, predicted to be deleterious and the gene has an association with patient's phenotype. In the case of metabolic disorders, novel variants that are non-LOF and additionally associated with no or inconclusive in vivo measurements of enzymatic activity and / or biomarker are classified as uncertain.

Likely benign: A likely benign variant is considered not likely to be the cause of the disease / phenotype. The main evaluation criteria refer to their frequency below 5% in general population, lack of observed impact on disease presence/severity/susceptibility, or non-segregation and/or co-occurrence detected. Classification as likely benign is additionally assigned to variants showing no damaging effect by in vivo measurements of enzymatic activity and / or biomarker levels.

Benign: A benign variant is not considered to be the cause of the disease/ phenotype. The main evaluation criteria refer to their frequency above 5% in general population, reported not to influence the disease risk of the individual, or predicted / shown to have no effect on protein or regulatory regions.