CENTOGENE
THE RARE DISEASE COMPANY

CentoLSD - Variant curation at CENTOGENE

PROCESS

Data gathering and variant curation procedures are developed and implemented in a web-based software which is compliant with the HUGO Gene Nomenclature Committee (HGNC; http://www.genenames.org), Human Genome Variant Society (HGVS; http://www.hgvs.org/mutnomen) and Human Phenotype Ontology (HPO; http://compbio.charite.de/phenomizer/) nomenclatures allowing collection of variants detected in the analyzed cases in nuclear coding, nuclear non-coding and mitochondrial genes. The software integrates in-house sample management systems and analysis platforms with external databases providing the curation scientist with a comprehensive and straightforward overview of all available and most up-to-date evidences. The curation process follows closely the American College of Medical Genetics (ACMG; https://www.ncbi.nlm.nih.gov/pubmed/25741868) guidelines, and considers particular adaptation for metabolic genes.

CURATION SCIENTISTS

Our curators are scientists with strong backgrounds in human genetics. They continuously undergo extensive training and competency testing to ensure curation consistency and standardization. They assure that data is error-free (including that items are properly associated and interpreted and that there are no inconsistencies or discrepancies against detected in-house observations and from external sources). They close the curation process by manual approval that reviewed and curated data comply with standard in-house procedures.

VARIANT CURATION WORKFLOW

Curation of variants starts with ascertaining reference sequences, standard ontology and nomenclature system. To maintain uniformity for variants naming, the curation scientists are following closely the HGVS nomenclature. The use of g. for genomic, c. for coding DNA seq, p. for protein and m. for mitochondrial is mandatory. The reference transcript is represented by longest transcript at amino acid level. Variants described on historical nomenclature are matched to current transcript/ nomenclature. At CENTOGENE, NCBI transcripts are used as standard reference transcript (RefSeq Gene) and the hg19 genomic build is considered. To standardize the use of gene symbols, the curation scientists follow the HGNC guidelines.

HGVS-compliant variants are then subjected to variant classification following the ACMG recommendations. Curation scientist evaluates all available criteria under the following categories: general population and sub-population allele frequencies and observations, computational and predictive tools, functional assays performed in house and / or obtained from literature, segregation

and de novo observations, annotations in literature and / or external and available databases (including disease-specific, locus-specific and variant databases), family history and allelic and genotyping data.

Variant-related information is predominantly extracted automatically, and when the case extracted / managed manually.
CENTOGENE has been 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.

At CENTOGENE variants are classified using five-tiered scheme: pathogenic, likely pathogenic, variant of uncertain significance (VUS), likely benign and benign. An additional clinical class, namely risk factor, represents a rare situation.
Variants are in general quarterly reviewed; however, if the system notifies curation scientists on new evidences to be available, the re-evaluation based on the new knowledge is prioritized and performed immediately.