CENTOGENE
THE RARE DISEASE COMPANY

CentoMDx 1.4

Handbook

Precautions/warnings:

For professional use only.

To support clinical diagnosis.

Not yet offered in the US.

Contents

Introduction

Diagnosing a patient with a rare disease is a complex task because not all existing genetic variants have been described or precisely annotated. Medical professionals need to obtain all available knowledge about a patient's detected genetic variants to establish the most accurate diagnosis possible.

CentoMDx™ is a device that compares the genetic information after sequencing analysis via an uploaded file with the dataset gathered from genetic tests conducted at CENTOGENE AG and, consequently, provides a diagnostic report with detailed information about the detected genetic findings. This means that every report generated by CentoMDx™ requires minimum effort and time from the user in order to provide the diagnosis based on the logic originated from and supported by extensive knowledge of the genetic variants in the 26 genes stored in CentoMD®.

This handbook gives details required to understand all steps of using CentoMDx™. It describes what information can be entered as patient details, what quality standards for uploaded file are required, how variant annotation is performed and what information can be found in the generated report. The accompanying CentoMDx™ user guide provides a detailed description of how to use this web based diagnostic service.

Intended use

CentoMDx™ is a browser-based software that is designed to provide a reliable recommendation of clinical diagnosis. Only pathogenic and likely pathogenic variants based on the dataset in CentoMD® will generate a clinical diagnostic report. The results in the report are intended to stand alone. The software will not allow independent review of data generated by the health care professional.

Facts and Features

CentoMDx™ provides a diagnostic service based on the expertise of CENTOGENE AG in the rare hereditary disease area and the acquired knowledge stored in CentoMD®. Pathogenic and likely pathogenic variants detected in CENTOGENE AG's unique population cohort serve as a dataset to annotate genetic findings after genetic testing. Evaluation of raw genetic data using CentoMDx™ is straightforward, the diagnostic statement is provided based on

the variant dataset without active decisions by a user. In only three steps – entering patient information, uploading the file and generating a report - CentoMDx™ provides the diagnostic report.

CentoMDx™ provides the following key features:

- Diagnostic service based on extensive knowledge of the genetic variants in 26 genes
- Option to enter patient information which will be displayed in the report
- Data file upload function
- Annotation of variants from an uploaded VCF file
- Report generation in a pdf format
- Clear diagnostic statement based on detected pathogenic and likely pathogenic variants
- Disease related findings summarized in a table
- Extended individual variant interpretation including software predictions, allele frequencies from various databases, reference to publications, related disease description and information about positive individuals analyzed at CENTOGENE AG
- Report storage and export functions
- User notification about reclassified reported variants

Patient details

CentoMDx™ provides the option to enter patient details that will be displayed in the diagnostic report. The following information can be included:

- Patient ID or Medical report number: Unique identifier assigned to each consented individual whose genetic data will be evaluated.

- Patient gender: Indicates the biological state of the individual of being male, female or other (e.g. intersex or unknown sex).

- Sample type: Type of sample sent for testing. Options include blood, tissue or other (e.g. DNA, dried blood spot or amniotic fluid).

- Biochemical analysis: Method to analyze enzymatic activity or levels of biomarkers in samples obtained from patients usually suspected of being affected by a metabolic disorder. This is a test performed via Tandem Mass Spectrometry to detect, diagnose, and monitor diseases, disease processes, and susceptibility, and to determine a course of treatment.
    - Enzyme (if available): Name of the enzyme encoded by the genes linked to metabolic diseases. Evaluation of the enzyme activity (if it was performed) compared to the reference interval can be entered as:
        - Normal: Levels of activity are within the normal range (no change).
        - Pathological: Levels of activity are significantly decreased compared to the normal range.

- Clinical information following HPO: Description of the features and characteristics of a patient entered as supporting evidence of the presence of a particular disease translated into the vocabulary defined by the HPO (http://www.human-phenotype-ontology.org/).

- Other relevant information: Information that is of importance for evaluation of a particular case (e.g. family history, consanguinity, etc.).

Data file upload

Based on CENTOGENE AG's extensive knowledge in the area of rare genetic disorders and the dataset of variants in CentoMD®, the user can initiate the evaluation of genetic data for patients and receive a diagnostic report using CentoMDx™.

Input file

The input file should be in Variant Call Format (VCF). CentoMDx™ supports VCF v.4.1 and later on the hg19 genome assembly. For specification, see https://samtools.github.io/hts-specs/VCFv4.1.pdf.

The input file must contain the 8 fixed mandatory columns ('#CHROM', 'POS', 'ID', 'REF', 'ALT', 'QUAL', 'FILTER', 'INFO'), followed by a 'FORMAT' column and then at least one column containing sample-specific genotype (GT) data. The genotype at every site is mandatory. If it is not found, the respective variant will be excluded. Acceptable formats are values separated by forward slash (e.g., 0/1) or pipe (e.g., 0|1). The sample in the 10th column is the "active sample". If the file contains additional columns with genotype data of additional samples or for the same sample from additional variant callers, they will be ignored.

The maximum number of annotated variants is 100,000. Once this limit is reached, any additional variants are excluded.

The following three parameters are required to meet CENTOGENE AG's quality criteria for diagnostics:

1. Site Quality (QUAL) should be >= 200.
2. Read depth (DP), if provided, should be >= 20 reads.
3. Allele frequency >= 30%. Note: allele frequencies will be calculated only if the file provides observation counts/depths for the alternate and reference alleles (AO/RO or AD).

Variants which fail to meet any of the above three parameters are excluded.

Information for variant annotation

CentoMD® variant information is used to annotate the genetic variants from an uploaded data file. The genetic diagnosis is based on detected class 1 (pathogenic) and class 2 (likely pathogenic) variants in 26 genes: ARSB, FUCA1, GAA, GALC, GALNS, GBA, GLA, GUSB, HEXA, HEXB, IDS, LIPA, MAN2B1, NAGLU, NIPA1, NPC1, NPC2, PKD1, PKD2, PKHD1, PPT1, SGCE, SGSH, SMPD1, SPG21, TSC1. For more information on these genes and the related diseases, see the table below (Table 1).

Gene symbol	Enzyme name	Associated disease	OMIM® disease	MOI
ARSB	Arylsulfatase B	Mucopolysaccharidosis type VI	253200	AR
FUCA1	Alpha-fucosidase	Fucosidosis	230000	AR
GAA	Acidic alpha-glucosidase	Pompe disease	232300	AR
GALC	Galactocerebrosidase	Krabbe disease	245200	AR
GALNS	N-acetylgalatosamine-6-sulfate-sulfatase	Mucopolysaccharidosis type IVA	253000	AR
GBA	Beta-glucocerebrosidase	Gaucher disease	230800	AR
GLA	Alpha-galactosidase	Fabry disease	301500	X-linked
GUSB	Beta-glucuronidase	Mucopolysaccharidosis type VII	253220	AR
HEXA	Beta-hexosaminidase A	Tay-Sachs disease	272800	AR
HEXB	Beta-hexosaminidase B	Sandhoff disease	268800	AR
IDS	Iduronate-2-sulfatase	Mucopolysaccharidosis type II	309900	X-linked
LIPA	Acid lipase	Wolman disease	278000	AR
MAN2B1	Alpha-mannosidase	Alpha-mannosidosis	248500	AR
NAGLU	N-acetyl-alpha-glucosaminidase	Mucopolysaccharidosis type IIIB	252920	AR
NIPA1	n.a.	Spastic paraplegia type 6	600363	AD
NPC1	n.a.	Niemann-Pick disease type C1	257220	AR
NPC2	n.a.	Niemann-Pick disease type C2	607625	AR
PKD1	n.a.	Polycystic kidney disease type 1	173900	AD
PKD2	n.a.	Polycystic kidney disease type 2	613095	AD
PKHD1	n.a.	Polycystic kidney disease	263200	AR
PPT1	Palmitoyl-protein thioesterase	Neuronal ceroid lipofuscinosis type 1	256730	AR
SGCE	n.a.	Myoclonus-dystonia	159900	AD
SGSH	n.a.	Mucopolysaccharidosis type IIIA	252900	AR
SMPD1	Acidic sphingomyelinase	Niemann-Pick disease type A	257200	AR
SPG21	n.a.	Spastic paraplegia type 21	248900	AR
TSC1	n.a.	Tuberous sclerosis type 1	191100	AD
FUCA1	Alpha-fucosidase	Fucosidosis	230000	AR

Table 1: Information on the 26 genes and related diseases for the CentoMDx™ diagnostic service is provided

Diagnostic report

The diagnostic report generated by CentoMDx™ provides information on the positive result when clinically relevant variants were detected. Besides the diagnostic statement disease, related findings are extensively described providing individual variant interpretation.

Diagnostic statement

The diagnostic statement in the report provides information about the presence of a disease based on the detected pathogenic and likely pathogenic variants in the 26 included genes taking into account the mode of inheritance of the disease as well as gender of the patient. The following diagnostic statements are possible:

- **A genetic diagnosis of (*disease name*) is confirmed.**
  - This statement is provided if:
    - a variant is detected in the case of an autosomal dominant mode of inheritance of the disease
    - for X-linked diseases either variant is detected in a male patient or in a female patient in a homozygous state
    - for autosomal recessive diseases, a variant is detected in a homozygous state.
- **The carrier status of (*disease name*) is genetically confirmed.**
  - This statement is provided if a variant is detected in heterozygous state for autosomal recessive disorders.
- **A genetic diagnosis of (*disease name*) is most likely confirmed.**
  - This statement is provided if more than one variant is detected in a heterozygous state for autosomal recessive disorders.
- **A/an/the (*gene symbol*) pathogenic variant(s) that can cause (*disease name*) has/have been detected.**
  - This statement is provided if variants are detected in a heterozygous state for a female patient in the case of an X-linked mode of inheritance of the disease.

Disease related findings

Disease related findings are summarized in the table where the following information is provided in the separate columns:

- **Gene**: A gene is defined by a sequence of DNA that represents a basic unit of heredity, being expressed in RNA and proteins. In CentoMDx™, the HGNC-approved gene symbols are used.

- **Transcript**: A digital nucleic acid sequence. Each gene is linked with a transcript or reference sequence. All variant-type annotations provide mapping to genomic coordinates (genome build hg19). Coding DNA reference sequence refers to a cDNA-derived sequence containing the full length of all coding regions and non-coding untranslated regions.

- **Change**: A sequence change at the coding DNA and protein levels. According to the reference sequence used, the genetic variants are linked with the corresponding location within the gene closely following the HGVS guidelines and recommendation.
  - Coding DNA change: Change at the cDNA level with numbering based on the coding DNA reference sequence.
  - Protein change: Change at the protein level with numbering based on the amino acid sequence, using one letter amino acid code and X designating a translation termination codon.

- **Zygosity**: Indicates whether a variant is detected on one chromosome or on both chromosomes and therefore describes the degree of similarity of the alleles for a trait in an organism.
  - Heterozygous: cells contain two different alleles of a gene at a locus.
  - Homozygous: identical alleles of the gene are present on both homologous chromosomes at a locus.
  - Hemizygous: Alleles detected in genes located on the X-chromosome for male cases.

Clinical significance according to CentoMD: Variant class in CentoMD® is based on the likelihood to predispose or to cause the observed phenotype/disease. The reported

genetic variants from the VCF file are either class 1 (pathogenic) or class 2 (likely pathogenic). The classification of genetic germline variants in CentoMD® is done according to the ACMG guidelines (Richards et al. (2015), Genet. Med., doi:10.1038/gim2015.30). Additionally, some modifications to the ACMG guidelines are applied. These modifications arise from our continuously growing internal expertise in the field of molecular diagnostics and are represented mainly by new evidences regarding internal observed frequencies, segregation data, genotype-phenotype correlation, co-occurrence, as well as enzymatic and biomarker levels. The adjustments to the ACMG recommendations are specified below.

Classification as pathogenic is additionally assigned to:

- Loss of function (LOF) variants that are associated with pathologically decreased biochemical levels/activities.
- Non-LOF variants that are associated with pathologically decreased biochemical levels/activities and where sufficient clinical information of the associated individual clearly supports the presence of the metabolic disease.

Classification as likely pathogenic is additionally assigned to:

- LOF-variants detected in the genes related to metabolic disorders with no biochemical evidences.
- Non-LOF-variants found in individual for whom pathological biochemical data is supporting but insufficient clinical information were provided to confirm the presence of the disease.

Variant re-evaluation and re-classification is a key feature of CentoMD® and performed regularly in the light of literature, publicly available clinical databases and most importantly, based on CENTOGENE AG's own continuously growing and improving proprietary information.

- **Disorder**: Name, Online Mendelian Inheritance in Man® (OMIM®) number and mode of inheritance (MOI) of the disease that is linked to the particular gene.
  Every disorder that has genetic component is described according to the OMIM® catalog. OMIM® was developed for the world-wide-web by NCBI and contains a list of human

genes and genetic diseases with links to other relevant resources (http://www.ncbi.nlm.nih.gov/omim). Every entry in OMIM® has a unique identifier, which is included in CentoMDx™. For example, OMIM® disease 230800 is Gaucher disease, type I. Additionally, each genetic disorder is linked with the observed mode of inheritance (MOI). MOI is defined as the manner in which a particular genetic trait or disorder is passed from one generation to the next. The following MOIs are included in CentoMDx™:

- Autosomal dominant (AD): The pattern of inheritance in which an affected individual has one copy of a mutant gene and one copy of normal gene on a pair of autosomal chromosomes.
- Autosomal recessive (AR): The pattern of inheritance in which both copies of an autosomal gene must be abnormal for a genetic condition or disease to occur.
- X-linked: The mode of inheritance of a trait encoded in the X chromosome.

Individual variant interpretation

Extended information is provided in the section of individual variant interpretation. This information is retrieved from various sources as described below.

- In silico predictions for the conservation of nucleotide and amino acid position as well as clinical significance of variants are supported by Alamut® Batch 1.8 software program.

- Variant frequencies expressed as a percentage are listed to provide an overview of the relative frequency of an allele at a particular locus in the general population. Information from different public databases such as 1000 Genomes, the Genome Aggregation Database (gnomAD) and the NHLBI GO Exome Sequencing Project (ESP) is included. The 1000 Genomes database provides a comprehensive description of common human genetic variation following whole-genome sequencing of a diverse set of individuals from multiple populations. gnomAD contains aggregated and harmonized exome and genome sequencing data from a variety of large-scale sequencing projects. The ESP database includes the datasets acquired by next-generation sequencing of the

protein coding regions of the human genome across diverse, richly phenotyped populations.

- If a genetic variant has previously been published in the literature, the main details of the publication are noted. For published variants the first author, issue date of the scientific paper and PubMed identifier (PMID) are indicated.

  Disease description provides a summary of information about the disorder caused by the variants in the particular gene. Summarized are characteristic signs of the disease, a spectrum of the clinical picture, differences of symptoms in female and male cases, common early and late symptoms, other signs, molecular relationship between genetic variation and phenotypic expression.
- Observations from individuals screened at CENTOTGENE AG are provided in the last paragraph of the individual variant interpretation. Variant frequency indicating the number of observations of the allele of interest at a particular locus in CENTOGENE AG's unique population cohort is shown. Additionally, information about positive individuals who carry pathogenic or likely pathogenic variants is included.

Glossary

Term	Explanation
Allele	One of two (or more) forms of a gene/genetic locus.
Allele frequency at CentoMD®	Indicates the number of observations of the allele of interest at a particular locus in CentoMD-unique population, expressed as decimal.
Biochemical analysis	Method to analyze enzymatic activity or levels of biomarkers in samples obtained from patients usually suspected being affected by a metabolic disorder. This is a test performed to detect, diagnose and monitor diseases, disease processes, susceptibility and determine a course of treatment.
Carrier	Individual who has only one copy of a genetic variant for a recessive disease.
Case	Indicates an individual where the diagnosis was confirmed by genetic testing.
cDNA	DNA that is synthesized from a messenger RNA template; the single-stranded form is often used as a probe in physical mapping.
cDNA change	Change at cDNA level following numbering based on coding DNA reference sequences.
Clinical information following HPO	Description of features and characteristics refereeing to the presence of a particular disease translated into the vocabulary defined by the HPO.
Clinical significance according to CentoMD®	Indicates the likelihood of this variant to predispose to or to cause the disorder.
Clinical significance-Likely pathogenic	Variant with probable pathogenicity, or the effect on the protein function is predicted to be likely deleterious (>90% probability to cause the disease).
Clinical significance-Pathogenic	Variant that is known to cause the phenotype/disease.
Disease	Particular abnormal, pathological condition that affects part or all of an organism. It is often construed as a medical condition associated with specific symptoms and signs.
Disease name	Name of a disease according to Online Mendelian Inheritance in Man® (OMIM®) database.
Enzyme interpretation	Evaluation of the enzyme activity compared to the reference interval.
Enzyme interpretation-Normal	Levels of enzyme activity are within the normal range (no change).
Enzyme interpretation-Pathological	Levels of enzyme activity are significantly decreased compared to the normal range.
Gene	Sequence of DNA that represents a basic unit of heredity, being expressed in RNA and proteins.
Gene symbol	A unique abbreviation for the gene name assigned by the HUGO Gene Nomenclature Committee (HGNC).
HGVS	Human Genome Variant Society that promotes: i) collection, documentation and distribution of genomic variation information and associated clinical variations; ii) guidelines and recommendations for mutation and gene nomenclature (http://www.hgvs.org/).
HGVS nomenclature	Standardized system recommended by HGVS to describe and document variant sequences.
HPO term	Phenotypic description of individuals provided by medical experts and translated into the vocabulary defined by the HPO.
Mode of Inheritance (MOI)	The manner in which a particular genetic trait or disorder is passed from one generation to the next.

Term	Explanation
MOI-Autosomal dominant (AD)	The pattern of inheritance in which an affected individual has one copy of a mutant gene and one normal gene on a pair of autosomal chromosomes.
MOI-Autosomal recessive (AR)	The pattern of inheritance in which both copies of an autosomal gene must be abnormal for a genetic condition or disease to occur.
MOI-X linked	The pattern of inheritance of a trait encoded on the X chromosome.
Mutation	Rare difference and permanent change in a DNA sequence or gene at a given locus. In medical genetics, it is often used to indicate a disease-causing allele.
OMIM®	Online Mendelian Inheritance in Man®: Database which contains a list of human genes and genetic diseases with links to other relevant sources, developed for the world-wide-web by NCBI (http://www.ncbi.nlm.nih.gov/omim).
OMIM® disease	Number of a disease according to Online Mendelian Inheritance in Man® (OMIM®) database.
Patient ID	Patient ID referring to a consented individual.
PMID	PubMed-Index for MEDLINE, PubMed identifier or PubMed unique identifier is a unique number assigned to each PubMed record.
Positive individual	Indicates an individual carrying a particular genetic variant.
Protein change	Change at protein level following numbering based on the amino acid sequence, using one letter amino acid code and X for designating a translation termination codon.
Sample type	Type of samples sent for testing.
Sample type-Blood	Blood sample sent for testing.
Sample type-Other	A sample sent to for testing which type is other than blood or tissue.
Sample type-Tissue	Tissue sample sent for testing.
Transcript	Digital nucleic acid sequence assembled by scientists as a representative example of a species' set of genes. Coding DNA reference sequence refers to a cDNA-derived sequence containing the full length of all coding regions and non-coding untranslated regions.
Variant	A sequence variation in a gene.
VCF	Variant Call Format: the format of a text file used in bioinformatics for storing gene sequence variations.
Zygosity	Indicates if a variant is detected on one chromosome or on both chromosomes. Describes the degree of similarity of the alleles for a trait in an organism.
Zygosity-Hemizygous	Used for alleles detected in genes located on X-chromosome for male cases.
Zygosity-Heterozygous	Gene locus when cells contain two different alleles of a gene.
Zygosity-Homozygous	Gene locus when identical alleles of the gene are present on both homologous chromosomes.