



CentoSLS

USER GUIDE

Contents

1. Purpose and Objective	3
2. How to access CentoLSD™	3
3. Understanding CentoLSD™ homepage	3
4. Selection of the gene of interest.....	5
5. Search and filter queries	5
6. Variant review.....	6
7. Glossary.....	8

1. Purpose and Objective

CentolSD™ is a software product designed, developed and maintained by CENTOGENE AG.

CentolSD™ is a comprehensive database that combines phenotype and genotype information gathered from samples of individuals analyzed at CENTOGENE AG. Every variant reported in CentolSD™ is linked to at least one clinically described case tested against Gaucher or Fabry disease through a validated and accredited laboratory workflow. CentolSD™ is a growing database; newly analyzed variants will be added quarterly.

This user guide has been designed to provide detailed instructions for the proper use of CentolSD™ in order to query, access and retrieve information from CENTOGENE AG's lysosomal storage disease (LSD) database. The following chapters provide step-by-step instructions for the use of CentolSD™. Additionally, you will find a glossary of terms and definitions in the attached appendix.

2. How to access CentolSD™

To access CentolSD™ go to <https://www.centogene.com/centolstd>. The database is freely available and no log-in is required.

3. Understanding CentolSD™ homepage

The primary CentolSD™ homepage is a results table –centric display that organizes information for all identified GBA (by default) genetic variants. On the top of the results table, you can download CentolSD™-related documents and find options to perform search queries (Figure 1).

Introducing CentoLSD™

In sharing our variant classification of *GBA* and *GLA* genetic changes, CENTOGENE is dedicated to full transparency and disclosure of our knowledge and understanding—critical for any clinical judgement.

CENTOGENE's comprehensive classification of variants is based on a highly qualified and standardized curation workflow. Our best-practice approach to curation workflow guarantees the best classification and the basis for clinical interpretation of newly identified variants. In addition, as part of our commitment to best practice we ensure that changes in variant classification will be proactively communicated by reclassification reports to all our past clinical cases as well.

We are confident in our decision statements concerning variant pathogenicity within CentoLSD™, which are all based on the combined experience of a dedicated team of geneticists, clinicians, and curation scientists.

Usage of CentoLSD™ is free, but subject to our terms and conditions, which can be found [here](#).

FURTHER INFORMATION – DOWNLOADS

- Gene-disease curation ▾
- Variant related documents ▾
- Curation by case ▾
- Consent for data sharing ▾
- User guide ▾

WORKING GROUPS

- Variant Interpretation Working Group ▾
- Curation Working Group ▾

Gene Selection → Select your gene of interest GBA GLA

Gene-related details →

Gene	Chromosome	Transcript	Disease	OMIM# disease	Inheritance
GBA	1	NM_0001573	Gaucher disease	230800	Autosomal recessive

Searching/ Filters → cDNA change Protein change gDNA change Location Coding effect Clinical significance ▾ Total variants for selected gene: 614

Variant Result Table

cDNA change	Protein change	gDNA change	Location	Coding effect	Clinical significance	Last review	Individuals [Het/Hom]	Internal allele frequency given at $\times 10^3$
c*245C>T		chr1.g.155204541G>A	3'UTR	Effect unknown	Uncertain	03/05/2020	10 [9/1]	0.00000
c*165T>C		chr1.g.155204621A>G	3'UTR	Effect unknown	Uncertain	04/04/2020	1467 [1359/108]	8.25954
c*102T>C		chr1.g.155204684A>G	3'UTR	Effect unknown	Likely benign	04/04/2020	232 [202/30]	0.14915
c*92G>A		chr1.g.155204694C>T	3'UTR	Effect unknown	Likely benign	03/30/2020	231 [219/12]	0.03728
c*29C>T		chr1.g.155204757G>A	3'UTR	Effect unknown	Uncertain	04/14/2020	33 [29/4]	0.21867

FIGURE 1 - Homepage organization

Terms linked to the symbol provide the corresponding definition. In order to understand the terminology simply press the symbol and a new window opens (see figure 2 as example). To close the window, you must click outside the definition display screen or the x symbol.

Select your gene of interest GBA GLA

Gene	Chromosome	Transcript	Disease
GBA	1	NM_0001573	Gaucher disease

cDNA change Protein change gDNA change Total entries for selected gene: 614

Internal allele frequency ✕

This term indicates the number of observations of the allele of interest at a particular locus in Centogene's unique population, expressed as decimal.

cDNA change	Protein change	gDNA change	Location	Coding effect	Clinical significance	Last review	Individuals [Het/Hom]	Internal allele frequency given at $\times 10^3$
c*245C>T		chr1.g.155204541G>A	3'UTR	Effect unknown	Uncertain	03/05/2020	10 [9/1]	0.00000
c*165T>C		chr1.g.155204621A>G	3'UTR	Effect unknown	Uncertain	04/04/2020	1467 [1359/108]	8.25954
c*102T>C		chr1.g.155204684A>G	3'UTR	Effect unknown	Benign	04/04/2020	232 [202/30]	0.14915
c*92G>A		chr1.g.155204694C>T	3'UTR	Effect unknown	Likely benign	03/30/2020	231 [219/12]	0.03728

FIGURE 2 - Terminology definition(s) review. Example: understanding the definition of the Internal allele frequency

The variant result table contains the following annotations: cDNA change, protein change (for coding variants), gDNA change, location, coding effect, clinical significance, last review, individuals and corresponding observed zygosity, and internal allele frequency (Figure 1; see Variant result table). Each row indicates a unique change at the DNA level in the gene of interest.

4. Selection of the gene of interest

The current version of CentoLSD™ contains two LSD- associated genes: GBA (glucosylceramidase beta) and GLA (galactosidase alpha). By default, the GBA genetic variants are displayed (Figure 3). You can retrieve GLA genetic variants by simply clicking on the GLA button.

Each activated gene contains details on chromosome location, transcript, disease, OMIM disease and inheritance.

The screenshot shows the 'Select your gene of interest' section. A red arrow points to the 'Pre-selected gene' label above the 'GBA' button. Another red arrow points to the 'GLA' button, which is labeled 'Available for selection'. Below this, a table provides details for the GBA gene:

Gene	Chromosome	Transcript	Disease	OMIM® disease	Inheritance
GBA	1	NM_0001573	Gaucher disease	230800	Autosomal recessive

Below the table are search and filter options: 'cDNA change', 'Protein change', 'gDNA change', 'Location', 'Coding effect', and 'Clinical significance'. A dropdown menu is visible next to 'Clinical significance'. On the right, it says 'Total entries for selected gene: 614'. At the bottom, there are icons for each search and filter option: cDNA change, Protein change, gDNA change, Location, Coding effect, Clinical significance, Last review, Individuals [Het/Hom], and Internal allele frequency given at $\times 10^3$.

FIGURE 3 - Gene selection

5. Search and filter queries

Above the results table, search and filtering options are provided. Search criteria refer to cDNA change, protein change, gDNA change and location; filters are included for coding effect and clinical significance (Figure 4).

Under the search boxes, type either a number or the exact variant. Example: for GBA gene you can search under protein change for 535 (to retrieve all protein changes at this codon) or p.R535C (to check only this specific genetic variant). While searching, the matches for the item of interest are indicated (see Figure 4a); you do not need to press Enter. To remove the searched criterion, click on the "x" symbol within search box and the system automatically refreshes.

Under filters, you can select one or more available items (see Figure 4b). To remove a set filter, simply de-select the item from the corresponding filter.

You can search for variants by using a combination of search boxes and / or filters. For example, by searching under GBA location for exon 12, activating the missense category under coding effect, and pathogenic and likely pathogenic under clinical significance, you will obtain all GBA missense variants classified as pathogenic and likely pathogenic within exon 12, as identified and classified at CENTOGENE.

a.

Select your gene of interest **GBA** **GLA**

Gene: GBA, Chromosome: 1, Transcript: NM_000157.3, Disease: Gaucher disease, OMIM® disease: 230800, Inheritance: Autosomal recessive

Search by Protein Change → cDNA change: 535, gDNA change: , Location: , Coding effect: , Clinical significance: - Total entries for selected gene: 614

Used search box is highlighted green

cDNA change	Protein change	gDNA change	Location	Coding effect	Clinical significance	Last review	Individuals [Het/Hom]	Internal allele frequency given at $\times 10^4$
c.1604G>C	p.R535P	chr1.g.155204793C>G	exon 12	Missense	Pathogenic	02/29/2020	1 [1/0]	0.00000
c.1604G>A	p.R535H	chr1.g.155204793C>T	exon 12	Missense	Pathogenic	02/29/2020	30 [29/1]	0.07043
c.1603C>T	p.R535C	chr1.g.155204794G>A	exon 12	Missense	Pathogenic	02/29/2020	10 [5/5]	0.00000

Number of Matches → 1 to 3 from 3 Results

b.

cDNA change: 535, gDNA change: , Location: , Coding effect: , Clinical significance: - Total entries for selected gene: 614

Filter indicated all items available for selection

cDNA change	Protein change	gDNA change	Location	Last review	Individuals [Het/Hom]	Internal allele frequency given at $\times 10^4$
c.1604G>C	p.R535P	chr1.g.155204793C>G	exon 12	02/29/2020	1 [1/0]	0.00000
c.1604G>A	p.R535H	chr1.g.155204793C>T	exon 12	02/29/2020	30 [29/1]	0.07043
c.1603C>T	p.R535C	chr1.g.155204794G>A	exon 12	02/29/2020	10 [5/5]	0.00000

1 to 3 from 3 Results

Effect unknown
 Frameshift
 In-frame
 Missense
 New translation termination codon
 Non-coding
 Nonsense
 Silent
 Splicing mutation

FIGURE 4 - How to use search and filter options **4a** - Searching under GBA / protein change by 535 **4b** - Additionally, filters can be applied (here coding effect filter is indicated)

6. Variant review

The variant of interest can be reviewed at two levels.

1. Variant annotations: every variant is linked to clinical class, last review date, number of individuals carrying this variant, its zygosity and the internal observed allele frequency. Example: the missense GBA variant p.R535C, is classified at CENTOGENE as pathogenic. This variant was identified in nine individuals; three individuals carried this variant in heterozygous state, six individuals carried this variant in homozygous state.
2. Variant rationale: every variant is linked to a detailed description of the evidences used to support the clinical significance. To read the variant rationale, click on the **i** symbol to left of the variant of interest. A new window opens where the summary can be read (Figure 5b).

a. Select your gene of interest GBA GLA

Gene: GBA Chromosome: 1 Transcript: NM_000157.3 Disease: Gaucher disease OMIM® disease: 230800 Inheritance: Autosomal recessive

cDNA change: p.R535C gDNA change: Location: Coding effect: Clinical significance: Total variants for selected gene: 614

cDNA change	Protein change	gDNA change	Location	Coding effect	Clinical significance	Last review	Individuals [Het/Hom]	Internal allele frequency given at $\times 10^{-4}$
c.1603C>T	p.R535C	chr1:g.155204794G>A	exon 12	Missense	Pathogenic	02/29/2020	10 [5/5]	0.14996

1 to 1 from 1 Results

b.

GBA c.1603C>T; p.R535C

HGVS: PVS2 PM2 PP2 PP3 PP4 PP5 PP6 PP7 PP8 PP9 PP10 PP11 PP12 PP13 PP14 PP15 PP16 PP17 PP18 PP19 PP20 PP21 PP22 PP23 PP24 PP25 PP26 PP27 PP28 PP29 PP30 PP31 PP32 PP33 PP34 PP35 PP36 PP37 PP38 PP39 PP40 PP41 PP42 PP43 PP44 PP45 PP46 PP47 PP48 PP49 PP50 PP51 PP52 PP53 PP54 PP55 PP56 PP57 PP58 PP59 PP60 PP61 PP62 PP63 PP64 PP65 PP66 PP67 PP68 PP69 PP70 PP71 PP72 PP73 PP74 PP75 PP76 PP77 PP78 PP79 PP80 PP81 PP82 PP83 PP84 PP85 PP86 PP87 PP88 PP89 PP90 PP91 PP92 PP93 PP94 PP95 PP96 PP97 PP98 PP99 PP100

Classified as: 02/29/2020 Assessed as: 04/07/2020

Rule ID: PVS2

Definition: Rare GBA variant confirming as CENPTOGENE a deleterious effect via in vivo measurements of Lyso-Gb1 biomarker (n=10 ng/ml; reference <=6.8 ng/ml).

Rule ID: PM2

Definition: Extremely low variant frequency for the autosomal recessive Gaucher disease in gnomAD and/or CensorHD (p<=0.005).

Rule ID: PP2

Definition: Missense variants on multiple NCBI transcripts (RefSeq), in a gene that has low rate of benign variants and in which missense are a common mechanism of the disease [min. 80% of the missense variants described in HGMD (D), ClinVar (consistently annotated pathogenic and likely pathogenic) and CensorHD are pathogenic and / or likely pathogenic].

Rule ID: PP5

Definition: Non-synonymous variants on multiple NCBI transcripts (RefSeq), where all in silico results (if any) must agree: SIFT (D), PolyPhen2_HDIV (D), PolyPhen2_HVAR (D), P1 Mutation taster (A, D), phyloP100way Vertebrate (n=15), GERP++ (n=2.0), CADD_raw (n=4.0), DANN score (n=0.5), RF_score (n=0.6), AdA_score (n=0.6).

Rule ID: PP6

Definition: Reputable source reports variant as pathogenic, evidence is not available for laboratory to perform an independent evaluation (HGMD - ClinVar consistently pathogenic and likely pathogenic; CensorHD pathogenic and likely pathogenic).

FIGURE 5 - Variant review; Example GBA, p.R535C **5a** - Variant annotations indicated by default **5b** - Rationale view

7. Glossary

TERM	EXPLANATION
Allele	One of two (or more) forms of a gene / genetic locus
Allele frequency	This term indicates the number of observations of the allele of interest at a particular locus in Centogene unique population, expressed as decimal x 10 ⁻³ .
Autosomal recessive	The pattern of inheritance in which both copies of an autosomal gene must be abnormal for a genetic condition or disease to occur.
cDNA change	Change at cDNA level following numbering based on coding DNA reference sequences.
Chromosome	A structure of nucleic acids and protein found in the nucleus of most living cells, carrying genetic information in the form of genes.
Clinical significance	Indicates the likelihood of this variant to predispose to or to cause the disorder.
Clinical significance- 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.
Clinical significance- 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.
Clinical significance- likely pathogenic	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 (>90% probability to cause the disease). 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.
Clinical significance- pathogenic	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.
Clinical significance- uncertain	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.
Coding effect	Describes the impact of the observed DNA change on protein level.

TERM	EXPLANATION
Coding effect- effect unknown	The coding effect on protein level has not been analyzed. An effect is expected but difficult to predict.
Coding effect- frameshift	A sequence change caused by deletion/insertion of nucleotides affecting an amino acid between the first (initiation, ATG) and last codon (termination, stop), replacing the normal C-terminal sequence with one encoded by another reading frame.
Coding effect- in frame	A sequence change that does not cause a shift in the triplet reading frame. As a result, one or more amino acids are replaced by one or more other amino acids.
Coding effect- missense	A single nucleotide change that results in a codon that codes for a different amino acid. Not all missense mutations are deleterious; some changes can have no effect. Because of the ambiguity of missense mutations, it is often difficult to interpret the consequences of these mutations in causing disease.
Coding effect- new translation termination codon	A sequence change that affects the translation termination codon (Ter/*) introducing a new downstream termination codon, extending the C-terminus of the encoded protein.
Coding effect- non coding	The change on DNA level that has no effect on protein or the effect of regulatory mutations is unknown.
Coding effect- nonsense	A sequence change that results in a premature stop codon, and in a truncated, incomplete protein product.
Coding effect- silent	A sequence change that results in a codon that codes for the same amino acid and without any functional change in the protein product.
Coding effect- splicing mutation	A sequence change that affects the splicing process (i.e. intron removal and exons joining). Splice-site mutations occur within genes in the noncoding regions (introns) just next to the coding regions (exons). Splice site mutations can eliminate an existing donor or acceptor site, which will cause an exon to be skipped and possibly result in a frameshift.
Coding effect- start loss	A sequence change in the ATG start codon that prevents the original start translation site from being used. This kind of mutation may eliminate gene function.
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.
gDNA change	Change at genomic DNA level following numbering based on genomic DNA reference sequence.
Gene	Sequence of DNA that represents a basic unit of heredity, being expressed in RNA and proteins.
Individual	It represents a unique individual tested for a certain disease, condition or carrier status at Centogene. Individual is expressed as number of individuals carrying the variant of interest and its zygosity (i.e. if variant is detected on one or on both chromosomes).
Individual- hemizygous	Hemizygous allele is an allele detected in genes located on X-chromosome for male cases.

TERM	EXPLANATION
Individual- heterozygous	Heterozygous is a gene locus when cells contain two different alleles of a gene.
Individual- homozygous	Homozygous is a gene locus when identical alleles of the gene are present on both homologous chromosomes.
Inheritance	The manner in which a particular genetic trait or disorder is passed from one generation to the next.
Location	The location of the DNA change relative to the transcriptional initiation site, initiation codon, polyadenylation site or termination codon of the corresponding gene.
OMIM	Online Mendelian Inheritance in Man: Database which contains a list of human genes and genetic diseases with links to other relevant resources, developed for the worldwide-web by NCBI (http://www.ncbi.nlm.nih.gov/omim).
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.
Transcript	The transcript that is used at CENTOGENE as a reference sequence.
X-linked	The pattern of inheritance of a trait encoded on the X chromosome.