CENTOGENE
THE RARE DISEASE COMPANY

CentoLSD

USER GUIDE

Contents

1. Purpose and Objective

CentoLSD™ is a software product designed, developed and maintained by CENTOGENE AG.
CentoLSD™ is a holistic 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 on use **https://www.centogene.com/centolsd**. The database is free avaible and no log in is required.

3. Understanding CentoLSD™ homepage

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

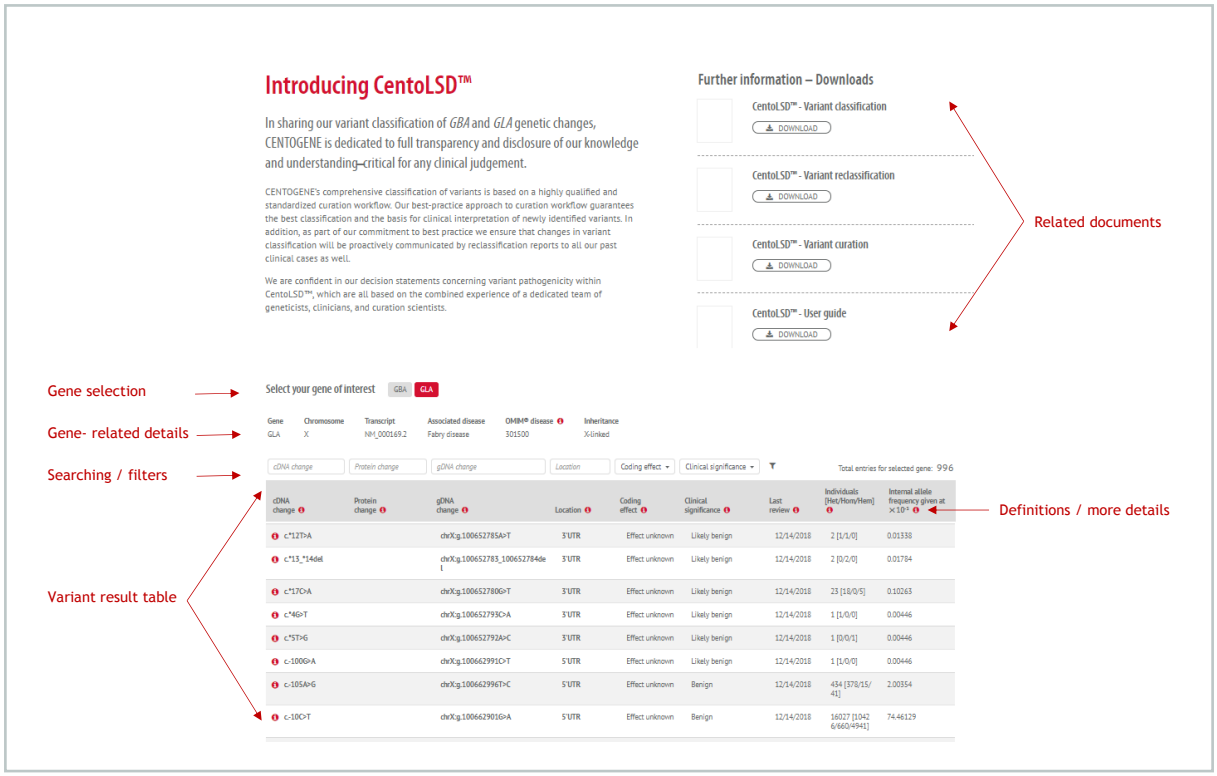

FIGURE 1 - Homepage organization

Terms linked to the symbol ⓘ provide their 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.

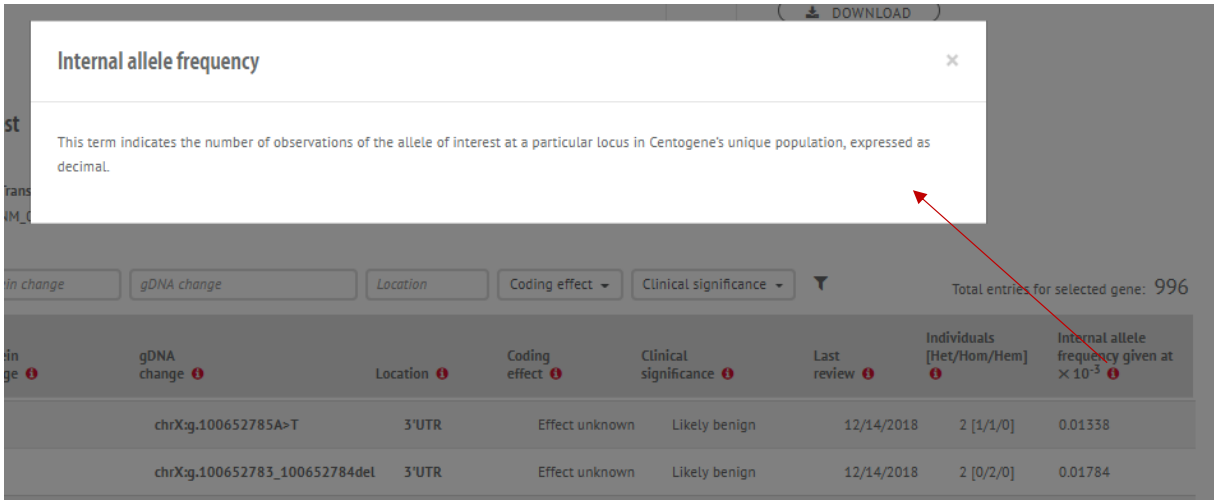

FIGURE 2 - Terminology definition(s) review.

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, internal allele frequency (Figure 1; see Variant result table). Each row indicates a unique change at DNA level within the gene of interest.

4. Selection of the gene of interest

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

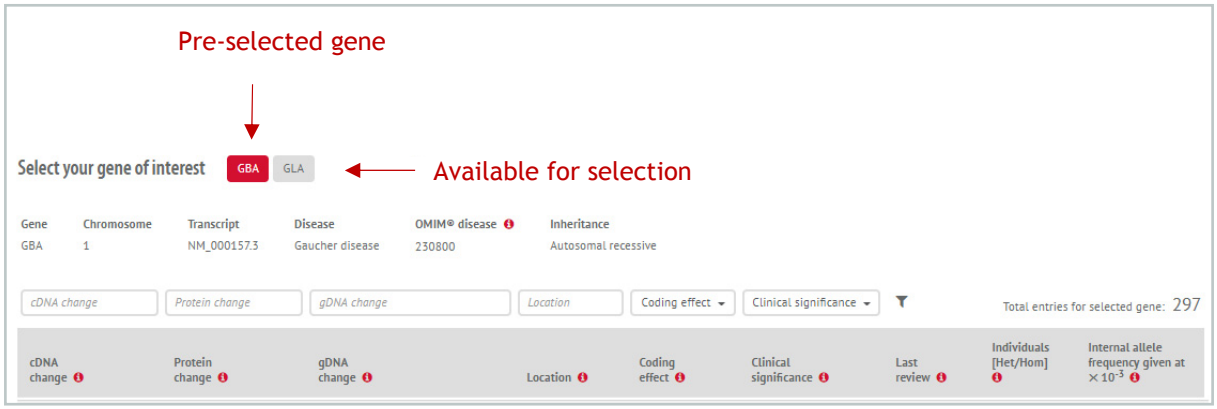

FIGURE 3 - Gene selection

5. Search and filter queries

Above the result table, searching 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 search boxes, type either a number or the exact searched 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 of 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 searching box. 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 under coding effect the missense category, and under clinical significance for pathogenic and likely pathogenic, you will obtained all GBA missense variants classified pathogenic and likely pathogenic within the exon 12, as identified and classified at CENTOGENE.

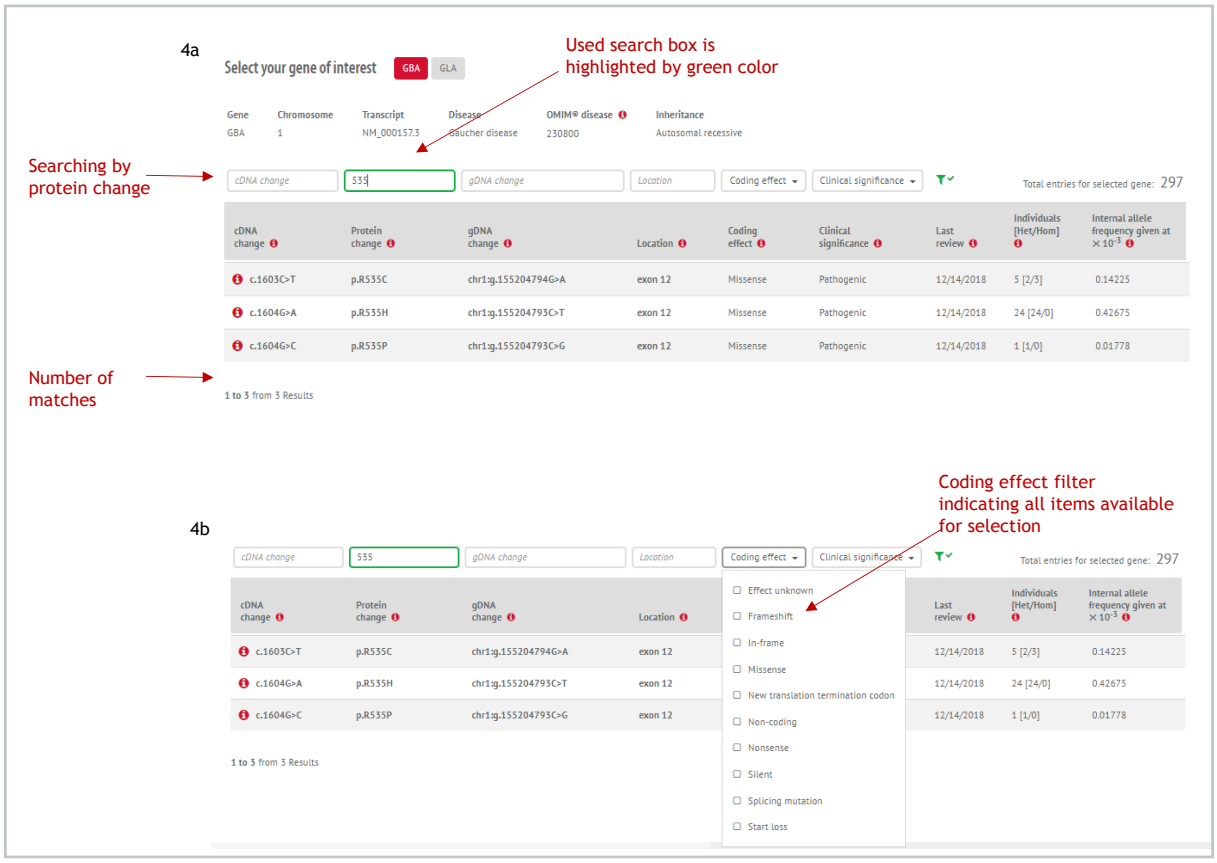

FIGURE 4 - How to use search and filter options

4a - Searching under GBA / protein change by 535

4b - Addionally, 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 five individuals; two individuals carried this variant in heterozygous state, three individuals carried this variant in homozygous state. The observed internal allele frequency is 0,14225 x 10^{-3} . See figure 5a for this example.
2. Variant rationale: every variant is linked to a detailed description of the observed evidences to support the clinical significance. To read the variant rationale, click on the ⓘ symbol attached to the variant of interest. A new window opens where the summary can be read. See figure 5b as example.

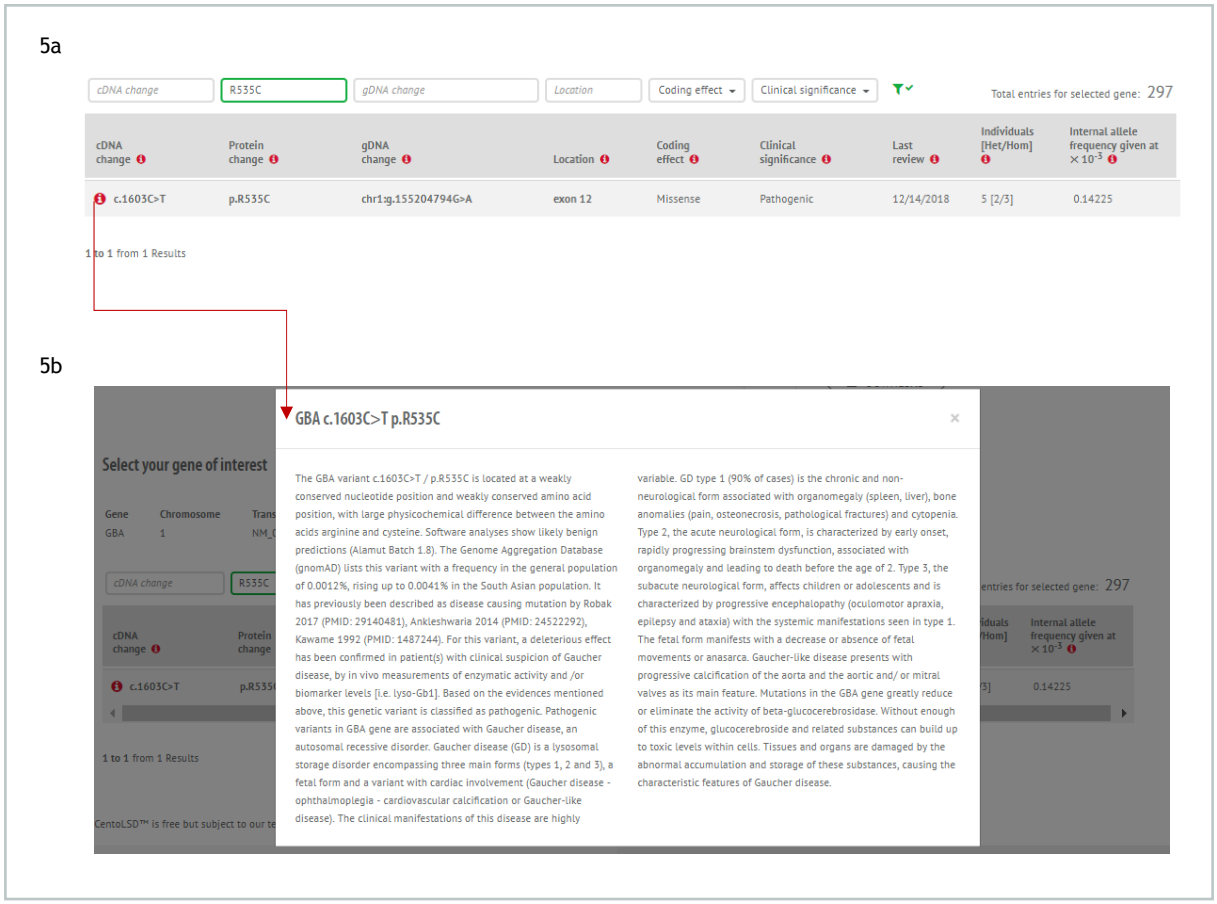

FIGURE 5 - Variant review; Example GBA, p.R535C

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.
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.
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 world-wide-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.