Human Mutation

OFFICIAL JOURNAL
HGVS
HUMAN GENOME VARIATION SOCIETY
www.hgvs.org

LOVD v.2.0: The Next Generation in Gene Variant Databases

Ivo F. A. C. Fokkema,[†] Peter E. M. Taschner,[*†] Gerard C. P. Schaafsma, J. Celli,
Jeroen F. J. Laros, and Johan T. den Dunnen

Center of Human and Clinical Genetics, Department of Human Genetics, Leiden University Medical Center, Leiden, Nederland

For the HVP Bioinformatics Special Issue

Received 12 October 2010; accepted revised manuscript 14 December 2010.
Published online 22 February 2011 in Wiley Online Library (www.wiley.com/humanmutation). DOI 10.1002/humu.21438

ABSTRACT: Locus-Specific DataBases (LSDBs) store information on gene sequence variation associated with human phenotypes and are frequently used as a reference by researchers and clinicians. We developed the Leiden Open-source Variation Database (LOVD) as a platform-independent Web-based LSDB-in-a-Box package. LOVD was designed to be easy to set up and maintain and follows the Human Genome Variation Society (HGVS) recommendations. Here we describe LOVD v.2.0, which adds enhanced flexibility and functionality and has the capacity to store sequence variants in multiple genes per patient. To reduce redundancy, patient and sequence variant data are stored in separate tables. Tables are linked to generate connections between sequence variant data for each gene and every patient. The dynamic structure allows database managers to add custom columns. The database structure supports fast queries and allows storage of sequence variants from high-throughput sequence analysis, as demonstrated by the X-chromosomal Mental Retardation LOVD installation. LOVD contains measures to ensure database security from unauthorized access. Currently, the LOVD Website (http://www.LOVD.nl/) lists 71 public LOVD installations hosting 3,294 gene variant databases with 199,000 variants in 84,000 patients. To promote LSDB standardization and thereby database interoperability, we offer free server space and help to establish an LSDB on our Leiden server.
Hum Mutat 32:557–563, 2011. © 2011 Wiley-Liss, Inc.

KEY WORDS: LSDB; database; open source; PHP; MySQL; LOVD

Introduction

The Leiden Open source Variation Database (LOVD), a freely available Web-based software for the collection, display, and curation of DNA variants in locus-specific databases (LSDBs) was developed following the LSDB-in-a-box concept [Fokkema et al., 2005]. The basic design criteria included a Web-based database system, based on freely available open source software, easy to use and install by noncomputer experts. Curators should be able to install a database system “out of the box” without complicated platform-specific programming or configuration steps. The LOVD system has been designed to comply with the guidelines and recommendations regarding content, design, and deployment of sequence variant databases developed by the HUGO Mutation Database Initiative, the Human Genome Variation Society (HGVS, http://www.hgvs.org/), and the Human Variome Project [Cotton et al., 2008; Scriver, 2000; Scriver et al., 1999].

The modular LOVD design provides the flexibility necessary to cope with the specific interests of the users of patient-centered, sequence variant-centered, disease-centered, and protein-centered databases. Since the original LOVD release, updates with new functional enhancements and bug fixes have been released regularly. Starting with the Leiden Muscular Dystrophy pages (http://www.DMD.nl/), many LSDBs have changed to an LOVD format. Because these transfers contribute to LSDB standardization and facilitate easy automated data retrieval, other LSDB curators are tempted to follow this example by using LOVD or other general software (UMD [Béroud et al., 2000] or Mutbase [Riikonen and Vihinen, 1999]). LOVD users have requested new features, for example, to allow curators to design different custom column sets to be used per gene database. To meet these requests, the database structure of LOVD v.2.0 has been redesigned without changing the original specifications [Fokkema et al., 2005]. This new design includes measures to prevent unauthorized data modification, to enhance data safety, and to repel Web attacks.

Materials and Methods

LOVD v.2.0 System Requirements, Database Installation, and Updates

The LOVD v.2.0 software is platform-independent. Use on a Web server requires the installation of the following freely available Open Source software: an HTTP Web server (e.g., Apache, http://www.apache.org), the PHP scripting language v.4.2.0 and up (http://www.php.net), and the MySQL database package v.3.23.33 and up (http://www.mysql.com). This software is installed on the servers of most (commercial) hosting providers. Ready-to-install binaries are available for most operating systems from the links provided above. The LOVD v.2.0 software is freely available from http://www.LOVD.nl. Detailed information about the installation, configuration, and upgrades is provided under the Documentation tab on the LOVD home page, which also provides access to the manual. Simple LOVD test installations (including Apache, PHP, and MySQL) can be set up on local Windows

Additional Supporting Information may be found in the online version of this article.

[†]Both authors contributed equally.

[*]Correspondence to: Peter E. M. Taschner, Department of Human Genetics S-4-P, Center for Human and Clinical Genetics, Leiden University Medical Center, Albinusdreef 2, P.O. Box 9600, 2300RC Leiden, Nederland. E-mail: P.Taschner@lumc.nl

Contract grant sponsor: The European Community's Seventh Framework Programme; Contract grant number: FP7/2007-2013, under grant agreement no. 200754—the GEN2PHEN project.

computers by downloading the CD image from the Website, burning the install CD, and inserting it into the computer. After following the instructions on the screen, LOVD can be set up in just a few minutes. Installing one of the regular updates only requires the old files to be overwritten on the server. The database backend will be upgraded automatically after the first activity by the database administrator, manager, or curator.

LOVD Extension Modules and Scripts

The modular design concept allows simple functionality extensions using specific modules. Three standard modules are prepackaged and installed automatically during LOVD v.2.0 installation and can be turned on and off on demand:

1. Mutalyzer nomenclature checker module—modifies the variant submission form to send the DNA change description and the reference sequence accession and version number to a remote server (http://www.mutalyzer.nl) running the new version 2 of the Mutalyzer software [Wildeman et al., 2008]. This tool performs an automatic check for compliance to the HGVS sequence variant nomenclature guidelines [Den Dunnen and Antonarakis, 2000; Den Dunnen and Paalman, 2003] to guarantee error-free entry of the sequence variants submitted. The Mutalyzer output containing the checked description and predicted effects at protein level will be displayed in a new window.
2. ShowMaxDBID—modifies the variant submission form for curators by presenting the first free value for the DB-ID field.
3. reCAPTCHA—modifies the submitter registration form to include an image displaying text that only humans can read. New submitters are required to type the words in the image to prevent fake registrations by spam bots.

Additional functionality is provided by other scripts. The GenBank File Uploader script allows curators to import custom reference sequences in GenBank format for checks with Mutalyzer and for use in the Reference Sequence Parser. The latter script helps curators to generate reference sequences in GenBank format with the appropriate transcript and protein annotation needed by Mutalyzer. The Reading Frame Checker uses information about the first and last exon deleted or duplicated in a gene to return the correct description of the change on DNA level as well as the predicted effect on protein level (in- or out-of-frame).

Database Structure

In the new database structure, patient data and sequence variant data are stored in separate tables to allow the creation of links between sequence variants in different genes and the same patient (see the LOVD v.2.0 database scheme in Supp. Fig. S1). This is particularly useful for oligogenic disorders, in which more than one gene has to be inactivated before manifestation, for example, Bardet-Biedl syndrome [Beales et al., 2003]. Data redundancy is reduced by storing the results of multiple mutation screens in one database which increases flexibility by allowing variant overviews per patient. System-wide views of all variant and patient information of a certain origin are supported if the patient geographic or ethnic origin columns have been enabled (see below).

Database Security

The LOVD software was designed to withstand all known types of attacks on Web-based database systems. These include: brute-forcing account passwords, SQL injection, XSS attacks, local and remote file inclusion, session hijacking and fixation, and network sniffing. Furthermore, the source code was checked using different security audit programs: CodeSecure^{TM} (Armorize Technologies, Santa Clara, CA) and Spike PHP Security Audit (http://developer.spikesource.com/projects/phpsecaudit), partially in co-operation with security specialists of the National Center for Biotechnology Information (NCBI). LOVD has two options to secure the personal curator, manager or database administrator accounts against password guessing. Each user can restrict access to their account to a certain IP address, an IP address range, or a list of these. Attempts to access the account from another computer will result in an error, even if the password is correct. Furthermore, an account will be locked by default after three failed log in attempts. Accounts can only be unlocked by a user of a higher level or, if enabled, by requesting LOVD to send a new password to the user's registered email address. LOVD logs all access attempts, including IP addresses from which they were made.

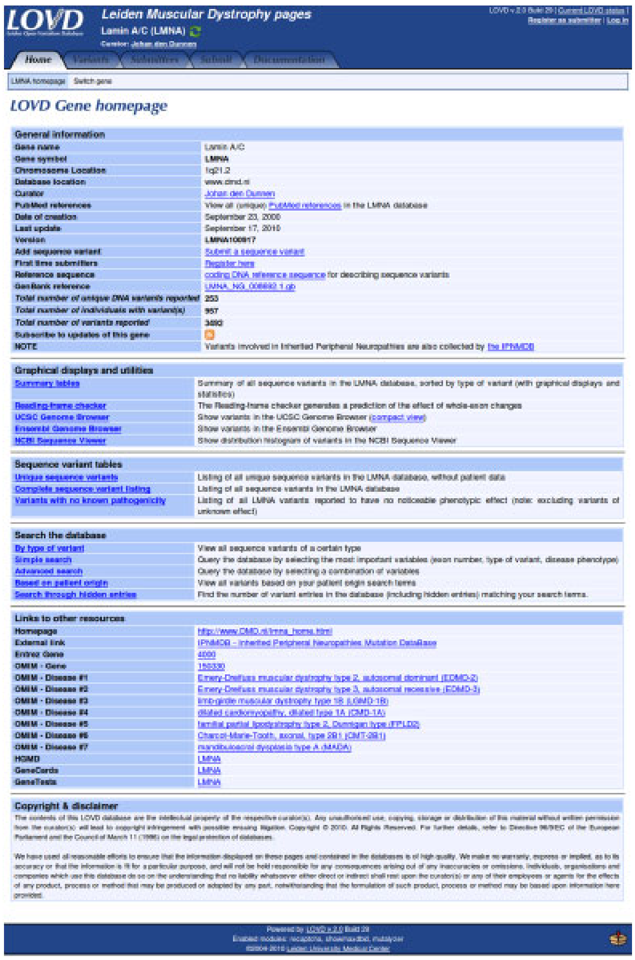
To protect against unauthorized modification of data, users need to confirm important data manipulations by entering their password before changes are saved. This prevents others, who use a computer where a user is logged in, from making modifications. If the server supports the secure socket layer (SSL) protocol for encryption, LOVD can be configured to force the use of SSL to secure the transfer of passwords or sensitive patient data over insecure networks. The transition from normal HTTP to an SSL connection will not be noticed when configured properly.

Results and Discussion

The LOVD User Interface

The homepage of a gene variant database displays general information about the gene and the database structure, and includes links to other sites of interest (e.g., OMIM, Entrez Gene, HGMD, etc.) as recommended by Claustres et al. [2002]. From here, users can access the allelic variant information and search the database. As shown with the Lamin A/C (LMNA) example from the Leiden Muscular Dystrophy pages (http://www.LOVD.nl/LMNA) a gene home page contains several new features (Fig. 1). The uppermost section of the page shows a green arrow, which allows users to select the gene of interest from the specific LOVD installation. To improve accreditation of the work performed, the curator's name is now displayed below the name of the gene. In compliance with the latest HGVS guidelines, the bottom of the homepage contains a copyright statement and a disclaimer. Tabs allow the user to search variants, view a list of submitters with contact information, submit new variants, and access the manual. Links in the top right corner provide access to information on the current status of the installation (e.g., the number of LSDBs on the server and the total number of sequence variants per database), submitter registration and the log in screen. A new feature of the general information section is a link to all PubMed references covered by the database entries.

The graphical display options contain links to the UCSC and Ensembl genome browsers and the NCBI sequence viewer, showing the stored variants in custom tracks. Summary tables show statistics on the percentages of variants per exon or the number of variants per bp per exon. This can reveal interesting information on the distribution of variants in the gene, such as specific hotspots. Unique for LOVD are separate diagrams showing the distribution and the numbers of the different variant types (substitutions, deletions, insertions, etc.) at the DNA, RNA, and protein levels, calculated automatically from the variant


LOVD
Leiden Muscular Dystrophy pages
Lamin A/C (LMNA)
Curator: Johan den Dunnen
LOVD v.2.0 Build 29 | Current LOVD status |
Register as submitter | Log in
Home
Variants
Submitters
Submit
Documentation
LMNA homepage
Switch gene

LOVD Gene homepage

General information	
Gene name	Lamin A/C
Gene symbol	LMNA
Chromosome Location	1q21.2
Database location	www.dmd.nl
Curator	Johan den Dunnen
PubMed references	View all (unique) PubMed references in the LMNA database
Date of creation	September 23, 2000
Last update	September 17, 2010
Version	LMNA100917
Add sequence variant	Submit a sequence variant
First time submitters	Register here
Reference sequence	coding DNA reference sequence for describing sequence variants
GenBank reference	LMNA_NG_008692.1.gb
Total number of unique DNA variants reported	253
Total number of individuals with variant(s)	957
Total number of variants reported	3490
Subscribe to updates of this gene	●
NOTE	Variants involved in Inherited Peripheral Neuropathies are also collected by the IPNMDB

Graphical displays and utilities	
Summary tables	Summary of all sequence variants in the LMNA database, sorted by type of variant (with graphical displays and statistics)
Reading-frame checker	The Reading-frame checker generates a prediction of the effect of whole-exon changes
UCSC Genome Browser	Show variants in the UCSC Genome Browser (compact view)
Ensembl Genome Browser	Show variants in the Ensembl Genome Browser
NCBI Sequence Viewer	Show distribution histogram of variants in the NCBI Sequence Viewer

Sequence variant tables	
Unique sequence variants	Listing of all unique sequence variants in the LMNA database, without patient data
Complete sequence variant listing	Listing of all sequence variants in the LMNA database
Variants with no known pathogenicity	Listing of all LMNA variants reported to have no noticeable phenotypic effect (note: excluding variants of unknown effect)

Search the database	
By type of variant	View all sequence variants of a certain type
Simple search	Query the database by selecting the most important variables (exon number, type of variant, disease phenotype)
Advanced search	Query the database by selecting a combination of variables
Based on patient origin	View all variants based on your patient origin search terms
Search through hidden entries	Find the number of variant entries in the database (including hidden entries) matching your search terms.

Links to other resources	
Homepage	http://www.DMD.nl/lmna_home.html
External link	IPNMDB - Inherited Peripheral Neuropathies Mutation DataBase
Entrez Gene	4000
OMIM - Gene	150330
OMIM - Disease #1	Emery-Dreifuss muscular dystrophy type 2, autosomal dominant (EDMD-2)
OMIM - Disease #2	Emery-Dreifuss muscular dystrophy type 3, autosomal recessive (EDMD-3)
OMIM - Disease #3	limb-girdle muscular dystrophy type 1B (LGMD-1B)
OMIM - Disease #4	dilated cardiomyopathy, dilated type 1A (CMD-1A)
OMIM - Disease #5	familial partial lipodystrophy type 2, Dunnigan type (FPLD2)
OMIM - Disease #6	Charcot-Marie-Tooth, axonal, type 2B1 (CMT-2B1)
OMIM - Disease #7	mandibuloacral dysplasia type A (MADA)
HGMD	LMNA
GeneCards	LMNA
GeneTests	LMNA

Copyright & disclaimer

The contents of this LOVD database are the intellectual property of the respective curator(s). Any unauthorised use, copying, storage or distribution of this material without written permission from the curator(s) will lead to copyright infringement with possible ensuing litigation. Copyright © 2010. All Rights Reserved. For further details, refer to Directive 96/9/EC of the European Parliament and the Council of March 11 (1996) on the legal protection of databases.

We have used all reasonable efforts to ensure that the information displayed on these pages and contained in the databases is of high quality. We make no warranty, express or implied, as to its accuracy or that the information is fit for a particular purpose, and will not be held responsible for any consequences arising out of any inaccuracies or omissions. Individuals, organisations and companies which use this database do so on the understanding that no liability whatsoever either direct or indirect shall rest upon the curator(s) or any of their employees or agents for the effects of any product, process or method that may be produced or adopted by any part, notwithstanding that the formulation of such product, process or method may be based upon information here provided.

Powered by LOVD v.2.0 Build 29
Enabled modules: recaptcha, showmaxdbs, mutalyzer
©2004-2010 Leiden University Medical Center

Figure 1. The LMNA database gene homepage. The underlined links on the Webpage provide access to the sequence variant information in the LMNA database or to external resources containing data about the LMNA gene and associated diseases. [Color figures can be viewed in the online issue, which is available at www.wiley.com/humanmutation.]

descriptions. The Reading-frame checker module is an additional feature, which can predict the effect of whole-exon changes on the reading frame (frame shifting or in frame). A news feed which automatically notifies subscribers of changes and additions to a specific gene or the complete database has been included.

Searching and Viewing Table Information

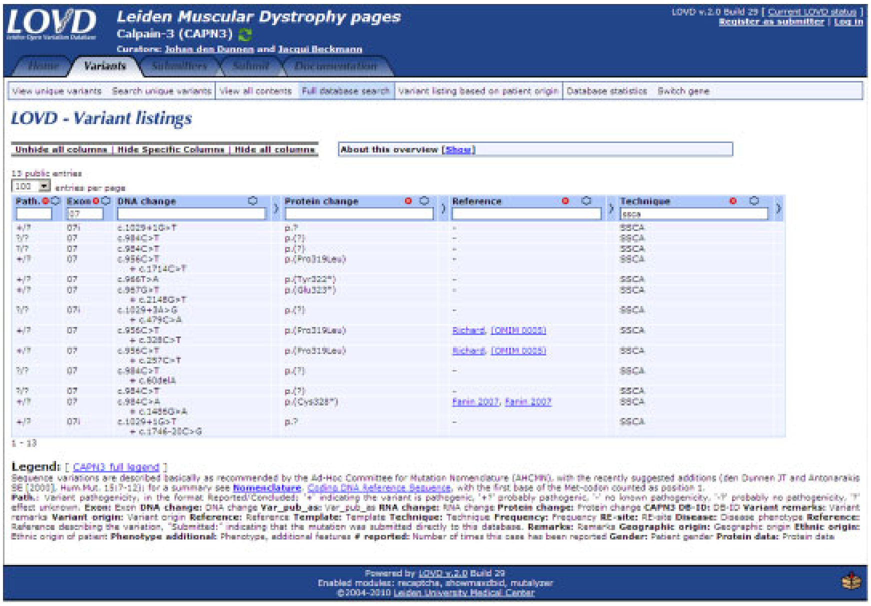
The "Search the database" section has been extended with two new options. The first is a search on the geographic or ethnic origin of the patient or the location of the laboratory reporting the variant. The latter feature allows collaborating labs to list all their contributions on their homepage, improving the accreditation received for the work performed. On the issue of country- or ethnicity-specific databases versus central databases, it should be noted that these additions diminish the need for separate databases. Separate databases will disperse data, reduce options to get a simple and up-to-date overview, and may lead to confusion if a patient is for example from a mixed background.

There are a variety of options to search and customize the user interface. LOVD can contain both public and nonpublic data. The second option "Search through hidden entries" can be enabled by curators to allow queries of restricted data. This search does not return the data itself, but the number of records matching the query. After clicking the Variants tab, the "Unique sequence variants" overview lists all the variant information recommended by the HGVS. Users can adapt the display to their preference by hiding columns and by sorting data on any column of interest (Fig. 2). Search boxes below the column headings support more detailed queries with Boolean operator symbols in any combination desired. Because the Boolean operators AND, OR, and NOT may be part of search strings, LOVD uses a space for AND, the pipe symbol | for OR and the exclamation mark ! for NOT. Boolean operator symbols should not be surrounded by spaces, for example, using Brazil|Canada in the Geographic Origin field of the LMNA database will return variants in patients from Brazil or Canada.

Published variants can be linked to PubMed using their PubMed ID, while the CustomLinks option facilitates direct linking to full text papers using their DOI. For variants present in OMIM [Hamosh et al., 2005], we provide a link in the Reference field of the entry. Variants present in dbSNP [Sherry et al., 2001] are entered as an independent record containing the DNA variant description, the predicted protein change, a link to dbSNP (Reference field) and the range of variant frequencies when reported in dbSNP (Frequency field). All dbSNP entries are linked (per gene) to one hypothetical patient called "dbSNP." This allows easy retrieval of dbSNP entries by searching the Reference field for "dbSNP" or just clicking one entry.

Data on variants found in animal models (including total gene knock outs and deletion or substitution variants) are also connected to a hypothetical patient (Patient-ID: "animal"). For example, mouse models carrying SGCA gene variants can be retrieved by searching the Variant remarks field for "mouse" (See http://www.lovd.nl/SGCA). The mouse variant is described at DNA, RNA, and protein level using the human reference sequence numbering. To clearly discriminate it from a true variant found in human, the description follows the HGVS rules for predicted effects and uncertain positions, which require the change to be shown in parentheses, for example, c.(1234G > T). The Remarks field clearly states the organism in which the variant was found; the Pathogenicity field is where functional consequences were observed. The same description format c.(1234G > T) is also used to list variants in pseudogenes that may be observed with nonspecific detection methods (e.g., nonunique PCR products). Again, the Remarks field is used to mention potential problems and to indicate the pseudogene. Searching for "c.(" will retrieve all these cases. Any available data on the functional analysis of gene variants are linked to a hypothetical patient (Patient-ID: "in vitro"). Each specific variant is described (DNA, RNA, Protein, Reference, etc.) with details of the test performed and the results obtained (Remarks field), connected to a classification of the variants functional consequences (Pathogenicity field).

Combination of variants are stored and displayed per patient (and per chromosome) (Fig. 3). To facilitate the interpretation of variant data in relation with disease, specific descriptions can be used to present information about the chromosome, which does not carry the variant. These descriptions are: c.= ("normal 2nd chromosome"), c.0 ("no paternal X-chromosome"), and c.? ("unknown variant in 2nd chromosome"). In patients affected by a recessive disease, one thus expects to identify one pathogenic variant for each of the two chromosomes. To highlight those cases where thus far variants in only one chromosome where detected, we add c.? for the other chromosome, with the remark field stating "unknown variant 2nd chromosome." For dominant diseases, the second chromosome is listed as c.=, with the remark "normal 2nd chromosome." In the case of X-linked recessive diseases, the second chromosome is listed as c.0 ("no paternal X-chromosome") in males and as c.= in females. We have noted that this increases the



Leiden Muscular Dystrophy pages
Calpain-3 (CAPN3)
Curators: Johan den Dunnen and Jacqui Beckmann

LOVD v.2.0 Build 29 [Current LOVD status]
Register as submitter | Log in

LOVD - Variant listings

13 public entries

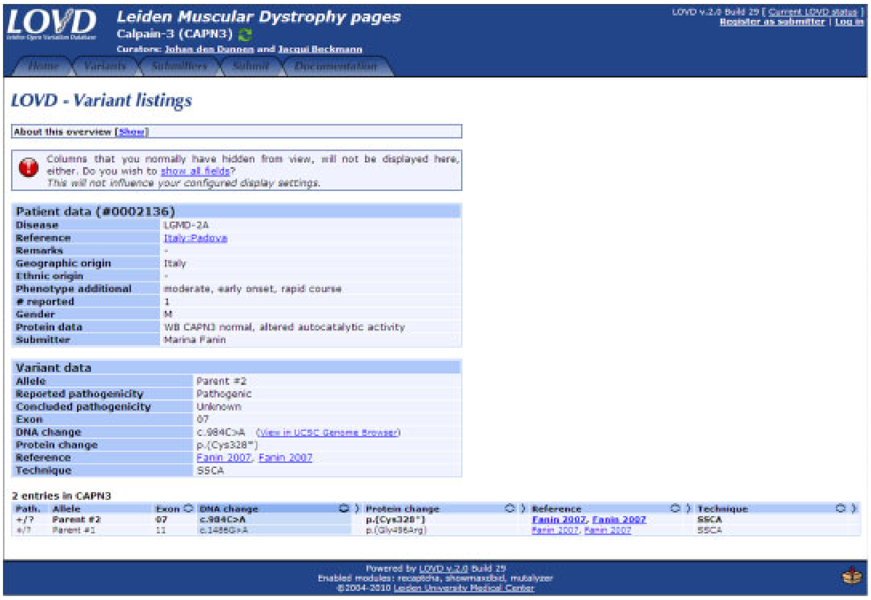
Path.	Exon	DNA change	Protein change	Reference	Technique
	07				ssca
+/?	07i	c.1029+1G>T	p.?	-	SSCA
?/?	07	c.984C>T	p.(?)	-	SSCA
?/?	07	c.984C>T	p.(?)	-	SSCA
+/?	07	c.956C>T + c.1714C>T	p.(Pro319Leu)	-	SSCA
+/?	07	c.966T>A	p.(Tyr322*)	-	SSCA
+/?	07	c.967G>T + c.2148G>T	p.(Glu323*)	-	SSCA
?/?	07i	c.1029+3A>G + c.479C>A	p.(?)	-	SSCA
+/?	07	c.956C>T + c.328C>T	p.(Pro319Leu)	Richard, [OMIM 0005]	SSCA
+/?	07	c.956C>T + c.257C>T	p.(Pro319Leu)	Richard, [OMIM 0005]	SSCA
?/?	07	c.984C>T + c.60delA	p.(?)	-	SSCA
?/?	07	c.984C>T	p.(?)	-	SSCA
+/?	07	c.984C>A + c.1486G>A	p.(Cys328*)	Fanin 2007, Fanin 2007	SSCA
+/?	07i	c.1029+1G>T + c.1746-20C>G	p.?	-	SSCA

1 - 13

Legend: [CAPN3 full legend]
Sequence variations are described basically as recommended by the Ad-Hoc Committee for Mutation Nomenclature (AHCMN), with the recently suggested additions (den Dunnen JT and Antonarakis SE [2000]. Hum.Mut. 15:7-12); for a summary see **Nomenclature**, Coding DNA Reference Sequence, with the first base of the Met-codon counted as position 1.
Path.: Variant pathogenicity, in the format Reported/Concluded: '+' indicating the variant is pathogenic, '+?' probably pathogenic, '-' no known pathogenicity, '-?' probably no pathogenicity, '?' effect unknown. **Exon:** Exon **DNA change:** DNA change **Var_pub_as:** Var_pub_as **RNA change:** RNA change **Protein change:** Protein change **CAPN3 DB-ID:** DB-ID **Variant remarks:** Variant remarks **Variant origin:** Variant origin **Reference:** Reference **Template:** Template **Technique:** Technique **Frequency:** Frequency **RE-site:** RE-site **Disease:** Disease phenotype **Reference:** Reference describing the variation, "Submitted:" indicating that the mutation was submitted directly to this database. **Remarks:** Remarks **Geographic origin:** Geographic origin **Ethnic origin:** Ethnic origin of patient **Phenotype additional:** Phenotype, additional features **# reported:** Number of times this case has been reported **Gender:** Patient gender **Protein data:** Protein data

Powered by LOVD v.2.0 Build 29
Enabled modules: recaptcha, showmaxdbid, mutalyzer
©2004-2010 Leiden University Medical Center

Figure 2. CAPN3 sequence variants displayed after a full database search for exon 7 allelic variants detected by single strand conformation analysis. The underlined links allow the user to view additional information about multiple database entries for the same variant or direct access to reference information in PubMed or OMIM. Several columns have been hidden from view. [Color figures can be viewed in the online issue, which is available at www.wiley.com/humanmutation.]



Leiden Muscular Dystrophy pages
Calpain-3 (CAPN3)
Curators: Johan den Dunnen and Jacqui Beckmann

LOVD v.2.0 Build 29 [Current LOVD status]
Register as submitter | Log in

LOVD - Variant listings

Columns that you normally have hidden from view, will not be displayed here, either. Do you wish to show all fields?
This will not influence your configured display settings.

Patient data (#0002136)	
Disease	LGMD-2A
Reference	Italy:Padova
Remarks	-
Geographic origin	Italy
Ethnic origin	-
Phenotype additional	moderate, early onset, rapid course
# reported	1
Gender	M
Protein data	WB CAPN3 normal, altered autocatalytic activity
Submitter	Marina Fanin

Variant data	
Allele	Parent #2
Reported pathogenicity	Pathogenic
Concluded pathogenicity	Unknown
Exon	07
DNA change	c.984C>A (View in UCSC Genome Browser)
Protein change	p.(Cys328*)
Reference	Fanin 2007, Fanin 2007
Technique	SSCA

2 entries in CAPN3

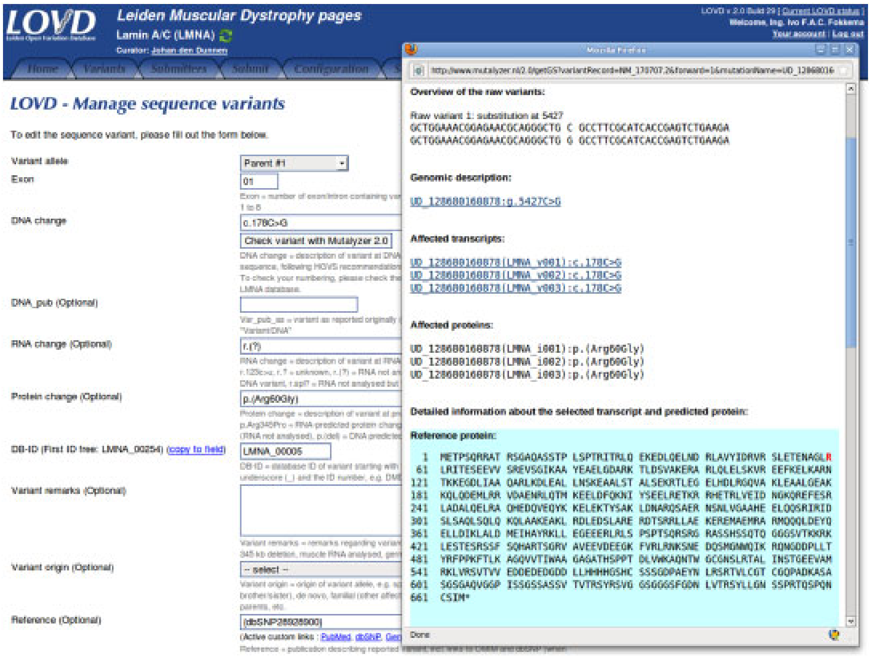
Path.	Allele	Exon	DNA change	Protein change	Reference	Technique
+/?	Parent #2	07	c.984C>A	p.(Cys328*)	Fanin 2007, Fanin 2007	SSCA
+/?	Parent #1	11	c.1486G>A	p.(Gly496Arg)	Fanin 2007, Fanin 2007	SSCA

Powered by LOVD v.2.0 Build 29
Enabled modules: recaptcha, showmaxdbid, mutalyzer
©2004-2010 Leiden University Medical Center

Figure 3. Patient data associated with CAPN3 variant c.984C>A. Data were accessed by clicking the c.984C>A information in Figure 2. The variant listing of the patient may contain additional variants in the same gene or in other genes in the same LOVD installation. [Color figures can be viewed in the online issue, which is available at www.wiley.com/humanmutation.]

awareness of nonexpert users, for example, carrier females in Duchenne Muscular Dystrophy, explaining why they remain unaffected despite carrying a nonsense variant in their genome. The c.=, c.0 and c.? variants receive database ID_00000 and are removed from the summary listings by the LOVD software.

Clicking on a variant will open a detailed view showing all entries carrying that variant, including patient and pathogenicity information. Columns can contain links to additional information (e.g., pedigree structure, growth curve, functional assay data, etc.). Access to this information can be unrestricted (public internet servers) or restricted (local servers or even the computer of the submitter only). One purpose of this option is to indicate that such information is present without making it publicly available. Selecting and clicking a specific row will open a detailed view showing all details per patient. Standard patient data fields include disease, geographic origin, ethnic origin, gender, and the name of the submitter. This overview also shows the details of the selected variant as well as a table with all other variants identified in the same and other genes of the same patient. Finally, graphical displays of the variants in several genome browsers (Ensembl [Flicek et al.,



LOVD Leiden Open Variation Database

Leiden Muscular Dystrophy pages
Lamin A/C (LMNA)
Curator: Johan den Dunnen

LOVD v.2.0 Build 29 [Current LOVD status]
Welcome, Ing. Ivo F.A.C. Fokkema
Your account | Log out

LOVD - Manage sequence variants

To edit the sequence variant, please fill out the form below.

Variant allele: Parent #1
Exon: 01
Exon = number of exon/intron containing va… 1 to 8
DNA change: c.178C>G
Check variant with Mutalyzer 2.0
DNA change = description of variant at DNA… sequence, following HGVS recommendation… To check your numbering, please check the… LMNA database.
DNA_pub (Optional):
Var_pub_as = variant as reported originally… "Variant/DNA"
RNA change (Optional): r.(?)
RNA change = description of variant at RNA… r.123c>u; r.? = unknown, r.(?) = RNA not an… DNA variant, r.spl? = RNA not analysed but…
Protein change (Optional): p.(Arg60Gly)
Protein change = description of variant at pr… p.Arg345Pro = RNA-predicted protein chang… (RNA not analysed), p.(del) = DNA predicted…
DB-ID (First ID free: LMNA_00254) (copy to field): LMNA_00005
DB-ID = database ID of variant starting with… underscore (_) and the ID number, e.g. DMD…
Variant remarks (Optional):
Variant remarks = remarks regarding varian… 345 kb deletion, muscle RNA analysed, gen…
Variant origin (Optional): -- select --
Variant origin = origin of variant allele, e.g. sp… brother/sister), de novo, familial (other affect… parents, etc.
Reference (Optional): {dbSNP28928900}
(Active custom links : PubMed, dbSNP, Gen…
Reference = publication describing reported variant, with links to OMIM and dbSNP (when…

Mozilla Firefox

http://www.mutalyzer.nl/2.0/getGS?variantRecord=NM_170707.2&forward=1&mutationName=UD_12868016

Overview of the raw variants:

```
Raw variant 1: substitution at 5427
GCTGGAAACGGAGAACGCAGGGCTG C GCCTTCGCATCACCGAGTCTGAAGA
GCTGGAAACGGAGAACGCAGGGCTG G GCCTTCGCATCACCGAGTCTGAAGA
```

Genomic description:

UD_128680168878:g.5427C>G

Affected transcripts:

UD_128680168878(LMNA_v001):c.178C>G
UD_128680168878(LMNA_v002):c.178C>G
UD_128680168878(LMNA_v003):c.178C>G

Affected proteins:

UD_128680168878(LMNA_i001):p.(Arg60Gly)
UD_128680168878(LMNA_i002):p.(Arg60Gly)
UD_128680168878(LMNA_i003):p.(Arg60Gly)

Detailed information about the selected transcript and predicted protein:

Reference protein:

```
  1  METPSQRRAT RSGAQASSTP LSPTRITRLQ EKEDLQELND RLAVYIDRVR SLETENAGLR
 61  LRITESEEVV SREVSGIKAA YEAELGDARK TLDSVAKERA RLQLELSKVR EEFKELKARN
121  TKKEGDLIAA QARLKDLEAL LNSKEAALST ALSEKRTLEG ELHDLRGQVA KLEAALGEAK
181  KQLQDEMLRR VDAENRLQTM KEELDFQKNI YSEELRETKR RHETRLVEID NGKQREFESR
241  LADALQELRA QHEDQVEQYK KELEKTYSAK LDNARQSAER NSNLVGAAHE ELQQSRIRID
301  SLSAQLSQLQ KQLAAKEAKL RDLEDSLARE RDTSRRLLAE KEREMAEMRA RMQQQLDEYQ
361  ELLDIKLALD MEIHAYRKLL EGEEERLRLS PSPTSQRSRG RASSHSSQTQ GGGSVTKKRK
421  LESTESRSSF SQHARTSGRV AVEEVDEEGK FVRLRNKSNE DQSMGNWQIK RQNGDDPLLT
481  YRFPPKFTLK AGQVVTIWAA GAGATHSPPT DLVWKAQNTW GCGNSLRTAL INSTGEEVAM
541  RKLVRSVTVV EDDEDEDGDD LLHHHHGSHC SSSGDPAEYN LRSRTVLCGT CGQPADKASA
601  SGSGAQVGGP ISSGSSASSV TVTRSYRSVG GSGGGSFGDN LVTRSYLLGN SSPRTQSPQN
661  CSIM*
```

Done

Figure 4. Submission of an LMNA variant checked via the Mutalyzer module. The results of the Mutalyzer 2 check are displayed in a new window. The protein variant description is shown between parentheses to indicate that it was derived by translation of the open reading frame of the reference sequence. [Color figures can be viewed in the online issue, which is available at www.wiley.com/humanmutation.]

2010], UCSC [Kent et al., 2002], NCBI Sequence Viewer (http://www.ncbi.nlm.nih.gov/projects/sviewer/)) are provided.

Sequence Variant Submission

The Web interface is publicly available and can be freely searched, but other activities, including sequence variant submission, require prior registration. Submitters can view, list, and edit all their records, add new patients or new variants to existing patients and generate personal submitted variant overviews to disseminate their work. When enabled, the sequence variant submission form supports checks of the variant description using the Mutalyzer module [Wildeman et al., 2008] (Fig. 4). When submitters add new information, the curators receive an e-mail copy as reminder and the entries will be nonpublic until approved by a curator.

Database Curation and Management

After curator approval, new variants are automatically displayed as public entries and their information is included in all linked Web pages. To support submission of larger datasets, curators can import tab-delimited data files provided by submitters or from previous LSDB versions. LOVD provides tools to quickly search, retrieve, and simultaneously change sets of entries. The search overview and simple pattern matching on specified fields or combinations of fields helps to select entries for editing. LOVD now includes a find and replace functionality, as well as a "copy column" feature facilitating simple database reformatting. More details about database customization by curators and database administrators can be found in Supp. Information S2.

Locus-Specific Databases Based Upon LOVD

Since the release and publication of LOVD [Fokkema et al., 2005], many curators have moved their database or static PDF and HTML files into an LOVD format. The use of LOVD has grown from the 17 mutation databases with 7,485 variants kept at the Leiden Muscular Dystrophy pages (http://www.DMD.nl/, February 18, 2005) to 71 public LOVD installations hosting 3,294 gene databases containing 199,000 variants in 84,000 patients (December 16, 2010). An overview of genes covered by public variant databases is available on the central LOVD server (see http://www.LOVD.nl for the current figures). Databases hosted on the Leiden servers are included by default. Database administrators who have set up their own online installation can activate the "share option" allowing communication with the central LOVD server. The data shared includes the name and location of the database, the names and number of the genes, the LOVD version, the number of (unique) variants and curator details. We have exploited this feature, together with the list of LSDBs provided by the HGVS, to create the URL http://geneID.lovd.nl (e.g., FKRP.lovd.nl), which allows users to be transferred directly to the database for the "geneID" gene (or a list of LSDBs when more than one exists).

With the support of the EU FP7-funded Gen2Phen project, we assist curators in the transfer of their LSDBs to a LOVD format. An example is the Mental Retardation database project (http://www.lovd.nl/MR), for which we created over 500 LSDBs for the X chromosome genes in a single installation in order to store the results of a large-scale resequencing study in patients with X-linked mental retardation [Tarpey et al., 2009]. Because individual access can be restricted to a subset of genes, large groups of curators can easily work together in a single installation without interfering with other users.

Display Tools

The standard installation comes with limited graphical display tools. This is due to the maximal flexibility we allow curators in extending LOVD with custom columns, making it difficult to automatically generate graphical displays. We have implemented methods to display data using the custom track option of the Ensembl and UCSC genome browsers and NCBI Sequence Viewer.

To accomplish this, LOVD stores chromosomal coordinates for all variants and uses this to transfer it to BED format. All variants will be shown relative to the annotation tracks in the genome browser. This allows the user to see local features, including known transcripts, conservation, dbSNP, structural variants, and flanking genes. In addition, users can use the tools provided by the genome browser to download (part) of the sequence, design PCR primers, etc. A similar display for individual variants is available by clicking the UCSC genome browser link in the detailed variant view.

Programmatic Queries of LOVD Installations

A new feature is an Application Programming Interface (API), which supports searches performed by computer programs. The API supports searching for genes based on gene symbol, chromosomal position or region and, per gene, for variants based on chromosomal positions or regions, positions or regions in the transcript, or DNA change description (see search formats in Supp. Information S3).

Data Sharing, Import, and Export

The flexibility achieved by creating custom columns is in itself detrimental for standardization. To bypass this, we have implemented the option to share custom column settings with another LOVD installation. To support standardization, column data can be exported and imported into other installations. Patient and variant data can also be easily imported and exported. LOVD supports multiple-gene data downloads for LSDB data sharing with central repositories according to the HGVS recommendations [Den Dunnen et al., 2009].

Further Developments

LOVD development is currently supported by the Gen2Phen project, funded by an EU FP7 grant. Within Gen2Phen, a general data model representing the data from genetic and genomic databases is under development with the ultimate goal being the development of a general data exchange format (see http://www.gen2phen.org/ for more information). Implementation of the Gen2Phen data model in database software will result in improved mapping of data between databases using the same syntax. This should increase their interoperability and improve automated data retrieval and submission from different sources. Because the data model includes all standard fields currently used in LOVD, UMD [Béroud et al., 2000] and Mutbase [Riikonen and Vihinen, 1999] database software, it should support data retrieval from these systems through world-wide queries once they are registered in the public gene variant database list. The LOVD query service can be regarded as a step toward accomplishing one of the aims of the Human Variome Project [Ring et al., 2006]: a central access point for the retrieval of variants in all genes worldwide, saving diagnostic laboratories, clinicians, and researchers considerable amounts of time and money [Cotton et al., 2008].

The Gen2Phen data model will be implemented in the database structure of LOVD v.3.0, which is currently under development (see http://www.lovd.nl/3.0/ for more information). In addition to gene, patient, and variant data objects, LOVD v.3.0 will contain new disease, phenotype, and screening data objects to support enhanced customization of phenotype and mutation screening information in gene variant databases within a single installation with more patient-centered, disease-centered, and protein-centered interests. Further improvement of interoperability will require semantic standardization of the database contents. To promote this, database software should support the use of ontologies and controlled vocabularies at least in fields containing information necessary for data exchange. The interoperability of genetic and genomic variant databases will also benefit from conversions of coding DNA positions used in LSDB variant descriptions to chromosomal positions used in genomic variant databases descriptions. Position conversions are part of the core functionality of the Mutalyzer nomenclature checker. LOVD already uses the new Mutalyzer 2 to calculate genomic positions for variant display in genome browsers. This interaction will be extended to handle position conversions in combination with a standard HGVS nomenclature check. In addition, future versions will use the reference sequence parsing capability of Mutalyzer 2.

Curators Needed

As part of the Gen2Phen project, we have recently generated a gene variant database for all genes linked to Mendelian disorders. Although we are working on importing data from literature, at the moment most of these databases are empty. Researchers with an informal database or even simple lists of gene variants are invited to contact us in order to upload these data. Similarly, this resource is suitable for storage of data obtained from exome/genome sequencing studies. In addition to gene variant data, these LSDBs need guardians: experts willing to devote some of their time to curate incoming data, promoting the database and uploading available data. If you are interested in extending your CV with an LSDB curatorship, please contact us. All feedback and efforts of the user community in extending the capabilities of LOVD by writing additional scripts are also warmly welcomed.

Acknowledgments

The authors specifically thank Raymond Dalgleish, and the many other LOVD database curators for their feedback and support, and Jean-Pierre Bayley for critically reading the manuscript.

References

Beales PL, Badano JL, Ross AJ, Ansley SJ, Hoskins BE, Kirsten B, Mein CA, Froguel P, Scambler PJ, Lewis RA, Lupski JR, Katsanis N. 2003. Genetic interaction of BBS1 mutations with alleles at other BBS loci can result in non-Mendelian Bardet-Biedl syndrome. Am J Hum Genet 72:1187–1199.

Béroud C, Collod-Béroud G, Boileau C, Soussi T, Junien C. 2000. UMD (universal mutation database): a generic software to build and analyze locus-specific databases. Hum Mut 15:86–94.

Claustres M, Horaitis O, Vanevski M, Cotton RGH. 2002. Time for a unified system of mutation description and reporting: a review of locus-specific mutation databases. Genome Res 12:680–688.

Cotton RG, Auerbach AD, Beckmann JS, Blumenfeld OO, Brookes AJ, Brown AF, Carrera P, Cox DW, Gottlieb B, Greenblatt MS, Hilbert P, Lehvaslaiho H, Liang P, Marsh S, Nebert DW, Povey S, Rossetti S, Scriver CR, Summar M, Tolan DR, Verma IC, Vihinen M, den Dunnen JT. 2008. Recommendations for locus-specific databases and their curation. Hum Mutat 29:2–5.

Den Dunnen JT, Antonarakis SE. 2000. Mutation nomenclature extensions and suggestions to describe complex mutations: a discussion. Hum Mutat 15:7–12.

Den Dunnen JT, Paalman MH. 2003. Standardizing mutation nomenclature: why bother? Hum Mutat 22:181–182.

Den Dunnen JT, Sijmons RH, Andersen PS, Vihinen M, Beckmann JS, Rossetti S, Talbot Jr CC, Hardison RC, Povey S, Cotton RG. 2009. Sharing data between LSDBs and central repositories. Hum Mutat 30:493–495.

Flicek P, Aken BL, Ballester B, Beal K, Bragin E, Brent S, Chen Y, Clapham P, Coates G, Fairley S, Fitzgerald S, Fernandez-Banet J, Gordon L, Gräf S, Haider S, Hammond M, Howe K, Jenkinson A, Johnson N, Kähäri A, Keefe D, Keenan S, Kinsella R, Kokocinski F, Koscielny G, Kulesha E, Lawson D, Longden I, Massingham T, McLaren W, Megy K, Overduin B, Pritchard B, Rios D, Ruffier M, Schuster M, Slater G, Smedley D, Spudich G, Tang YA, Trevanion S, Vilella A, Vogel J, White S, Wilder SP, Zadissa A, Birney E, Cunningham F, Dunham I, Durbin R, Fernández-Suarez XM, Herrero J, Hubbard TJ, Parker A, Proctor G, Smith J, Searle SM. 2010. Ensembl's 10th year. Nucleic Acids Res 38: D557–D562.

Fokkema IF, den Dunnen JT, Taschner PE. 2005. LOVD: easy creation of a locus-specific sequence variation database using an "LSDB-in-a-Box" approach. Hum Mutat 26:63–68.

Hamosh A, Scott AF, Amberger JS, Bocchini CA, McKusick VA. 2005. Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. Nucleic Acids Res 33:D514–D517.

Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D. 2002. The human genome browser at UCSC. Genome Res 12:996.

Riikonen P, Vihinen M. 1999. MUTbase: maintenance and analysis of distributed mutation databases. Bioinformatics 15:852–859.

Ring HZ, Kwok PY, Cotton RG. 2006. Human Variome Project: an international collaboration to catalogue human genetic variation. Pharmacogenomics 7:969–972.

Scriver CR. 2000. Guidelines and recommendations for content, structure, and deployment of mutation databases: II. Journey in progress. Hum Mutat 15:13–15.

Scriver CR, Nowacki PM, Lehväslaiho H. 1999. Guidelines and recommendations for content, structure, and deployment of mutation databases. Hum Mutat 13: 344–350.

Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, Sirotkin K. 2001. dbSNP: the NCBI database of genetic variation. Nucleic Acids Res 29: 308–311.

Tarpey PS, Smith R, Pleasance E, Whibley A, Edkins S, Hardy C, O'Meara S, Latimer C, Dicks E, Menzies A, Stephens P, Blow M, Greenman C, Xue Y, Tyler-Smith C, Thompson D, Gray K, Andrews J, Barthorpe S, Buck G, Cole J, Dunmore R, Jones D, Maddison M, Mironenko T, Turner R, Turrell K, Varian J, West S, Widaa S, Wray P, Teague J, Butler A, Jenkinson A, Jia M, Richardson D, Shepherd R, Wooster R, Tejada MI, Martinez F, Carvill G, Goliath R, de Brouwer AP, van Bokhoven H, Van Esch H, Chelly J, Raynaud M, Ropers HH, Abidi FE, Srivastava AK, Cox J, Luo Y, Mallya U, Moon J, Parnau J, Mohammed S, Tolmie JL, Shoubridge C, Corbett M, Gardner A, Haan E, Rujirabanjerd S, Shaw M, Vandeleur L, Fullston T, Easton DF, Boyle J, Partington M, Hackett A, Field M, Skinner C, Stevenson RE, Bobrow M, Turner G, Schwartz CE, Gecz J, Raymond FL, Futreal PA, Stratton MR. 2009. A systematic, large-scale resequencing screen of X-chromosome coding exons in mental retardation. Nat Genet 41:535–543.

Wildeman M, van Ophuizen E, den Dunnen JT, Taschner PE. 2008. Improving sequence variation descriptions in mutation databases and literature using the Mutalyzer mutation nomenclature checker. Hum Mutat 29:6–13.