



www.gen2phen.org

D4.1 User-Manual for LSDB-in-a-box V1 Software

WP4 – Genetics G2P Databases

**V4.0
Final**

Lead beneficiary: LUMC

Date: 27/05/2009

Nature: Report

Dissemination level: PU



 HEALTH-200754	D4.1 User-Manual for LSDB-in-a-box V1 Software		
	WP4 - Genetics G2P databases		Security: PU
	Author(s): P. Taschner (LUMC), I. Fokkema (LUMC), C. Beroud (INSERM), J. den Dunnen (LUMC)		Version: v4.0–Final

TABLE OF CONTENTS

DOCUMENT INFORMATION	3
DOCUMENT HISTORY	3
DEFINITIONS	4
1. INTRODUCTION.....	5
2. DESCRIPTION OF WORK	5
3. FUTURE WORK.....	5
ANNEX	8
REFERENCES.....	9

 HEALTH-200754	D4.1 User-Manual for LSDB-in-a-box V1 Software		
	WP4 - Genetics G2P databases		Security: PU
	Author(s): P. Taschner (LUMC), I. Fokkema (LUMC), C. Beroud (INSERM), J. den Dunnen (LUMC)		Version: v4.0–Final

Document Information

Grant Agreement Number	HEALTH-F4-2007-200754	Acronym	GEN2PHEN
Full title	Genotype-To-Phenotype Databases: A Holistic Solution		
Project URL	http://www.gen2phen.org		
EU Project officer	Frederick Marcus (Frederick.Marcus@ec.europa.eu)		


Deliverable	Number	D4.1	Title	User-Manual for LSDB-in-a-box V1 Software
Work package	Number	4	Title	WP4 - Genetics G2P Databases

Delivery date	Contractual	June 2009	Actual	August 2009
Status	Final Draft		final <input checked="" type="checkbox"/>	
Nature	Report <input checked="" type="checkbox"/> Prototype <input type="checkbox"/> Other <input type="checkbox"/>			
Dissemination Level	Public <input checked="" type="checkbox"/> Confidential <input type="checkbox"/>			

Authors (Partner)	P. Taschner (LUMC), I. Fokkema (LUMC), C. Beroud (INSERM), J. den Dunnen (LUMC)			
Responsible Author	Johan T. den Dunnen		Email	J.T.den_Dunnen@lumc.nl
	Partner	LUMC	Phone	+31-71-5269501

Document History

Name	Date	Version	Description
P. Taschner, I. Fokkema, J. den Dunnen	27/04/09	1.0	Draft
P. Taschner	11/05/09	2.0	Revision
P. Taschner	26/05/09	3.0	Revision
P. Taschner, C. Beroud	27-5-09	4.0	Final draft

 HEALTH-200754	D4.1 User-Manual for LSDB-in-a-box V1 Software		
	WP4 - Genetics G2P databases	Security: PU	
	Author(s): P. Taschner (LUMC), I. Fokkema (LUMC), C. Beroud (INSERM), J. den Dunnen (LUMC)	Version: v4.0–Final	4/9

Definitions

- Partners of the GEN2PHEN Consortium are referred to herein according to the following codes:

ULEIC – University of Leicester (UK) – Coordinator

EMBL – European Molecular Biology Laboratory (Germany) – Beneficiary

FIMIM – Fundació IMIM (Spain) – Beneficiary

LUMC – Leiden University Medical Center (Netherlands) – Beneficiary

INSERM – Institut National de la Santé et de la Recherche Médicale (France) – Beneficiary

KI – Karolinska Institutet (Sweden) – Beneficiary

FORTH – Foundation for Research and Technology Hellas (Greece) – Beneficiary

CEA – Commissariat à l’Energie Atomique (France) – Beneficiary

EMC – Erasmus Universitair Medisch Centrum Rotterdam (Netherlands) – Beneficiary

UH.FGC – Helsingin Yliopisto (Finland) – Beneficiary

UAVR – Universidade de Aveiro (Portugal) – Beneficiary

UWC – University of the Western Cape (South Africa) – Beneficiary

CSIR – Council of Scientific and Industrial Research (India) – Beneficiary

SIB – Swiss Institute of Bioinformatics (Switzerland) – Beneficiary

UNIMAN – The University of Manchester (UK) – Beneficiary


BIOBASE – BioBase GmbH. (Germany) – Beneficiary

deCODE – Islensk Erfoagreining EH (Iceland) – Beneficiary

PHENO – Phenosystems S.A. (Belgium) – Beneficiary

BCP – Biocomputing Platforms Ltd. Oy (Finland) – Beneficiary

- Grant Agreement:** The agreement signed between the beneficiaries and the European Commission for the undertaking of the GEN2PHEN project (HEALTH-200754).
- Project:** The sum of all activities carried out in the framework of the Grant Agreement by the Consortium.
- Work plan:** Schedule of tasks, deliverables, efforts, dates and responsibilities corresponding to the work to be carried out for the GEN2PHEN project, as specified in Annex I to the Grant Agreement.
- Consortium:** The GEN2PHEN Consortium, conformed by the above-mentioned legal entities.
- Consortium agreement:** agreement concluded amongst GEN2PHEN participants for the implementation of the Grant Agreement. Such an agreement shall not affect the parties’ obligations to the Community and/or to one another arising from the Grant Agreement.

 HEALTH-200754	D4.1 User-Manual for LSDB-in-a-box V1 Software		
	WP4 - Genetics G2P databases		Security: PU
	Author(s): P. Taschner (LUMC), I. Fokkema (LUMC), C. Beroud (INSERM), J. den Dunnen (LUMC)		Version: v4.0–Final


1. INTRODUCTION

Databases storing DNA sequence variation information exist in two types: a) comprehensive (genomic) databases, which contain information about all genes (e.g. Online Mendelian Inheritance in Man (OMIM, 1), Single Nucleotide Polymorphism database (dbSNP, 2)) and b) Locus-specific (or specialized) databases (LSDBs) covering one or more specific genes. The latter have been developed by curators with different interests, leading to patient-centered, sequence variation-centered, disease-centered, and protein-centered databases. LSDB examples can be found through the Variation Databases page of the Human Genome Variation Society (HGVS, www.HGVS.org/).

The rationale behind Work package 4 “Genetics G2P Databases” within the Gen2Phen project is that the complete collection and expert curation of gene sequence variants in these databases and their coupling to phenotypic consequences (if any) will be essential for proper future healthcare and research. WP4 builds upon the community consultation and minimum agreed content standard determined by WP2 and the data models developed within WP3. WP4 aims to contribute standardized technologies and modular components to meet the needs of individual researchers and clinical geneticists interested in very detailed information relating to single genes or single diseases. The work in this package is focussed on creating locus-specific database solutions, including one or more self-contained ‘LSDB-in-a-box’ applications. The LSDB-in-a-box concept means that LSDB curators should be able to install the LSDB system "out of the box" without complicated platform-specific programming or configuration steps. Furthermore, the modular design of the database system should provide the flexibility to cope with the different interests of the curators. In addition, WP4 provides server capacity to host LSDBs for database projects that lack adequate local IT hardware and support.

2. DESCRIPTION OF WORK

GEN2PHEN partners INSERM and LUMC already have legacy implementations of ‘LSDB-in-a-box’ software: the Universal Mutation Database (UMD - <http://www.umd.be/>) and the Leiden Open-source Variation Database (LOVD- <http://www.lovd.nl>)(3,4). Both applications support federated LSDBs as suggested at the Human Variome Project meeting (5). They can store DNA sequence variation information data generated by different clinicians and researchers for one or more specific genes in combination with phenotypic and disease-related information. Clinicians and researchers can search the database to assess the potential disease-causing effects of sequence variants in specific genes or to see which diseases or phenotypes are associated with variants in different genes.

 HEALTH-200754	D4.1 User-Manual for LSDB-in-a-box V1 Software		
	WP4 - Genetics G2P databases		Security: PU
	Author(s): P. Taschner (LUMC), I. Fokkema (LUMC), C. Beroud (INSERM), J. den Dunnen (LUMC)		Version: v4.0–Final

UMD has a strong position both for researchers and clinicians with the inclusion of many phenotype-specific tools. It can also be used as a local database system in DNA diagnostic laboratories. UMD software for Windows 98, 2000 and NT operating systems and Mac OS 9 and OS X operating systems can be downloaded freely from the UMD website (<http://www.umd.be/>).


UMD features:

- Automated mutation nomenclature system regularly updated to follow HGVS nomenclature,
- Single system to develop a local database and a web server.
- Support various levels of access through a login/password system.
- Full search system (combination of multiple criteria from different fields).
- Many analysis tools (UMD is a knowledgebase system):
- Evaluation of consequences of a variation using new algorithms such as the “UMD predictor” to predict the pathogenicity of substitutions (6) and/or the impact on mRNA (7).
- Search for repeated motifs involved in mutational events, tables or graphical displays of mutations’ profiles, mutational events, geographic distribution of mutations.
- Clinical tools that can handle unlimited numbers of symptoms, select subgroups and perform statistical analysis to specify natural history of the diseases, establish genotype-phenotype correlations and much more.
- Tools for new genotype-based therapies (exon-skipping (6), nonsense readthrough...)
- Supports data exchange with central systems or other databases.

LOVD has a strong position as a mainly web-based system populated by genetics researchers. The platform-independent LOVD software can be downloaded freely from the LOVD website (<http://www.lovd.nl>). For local installation of LOVD on a Windows PC, a CD is available for download from the LOVD website. Using LOVD’s simple installation scripts, an LSDB following the HGVS standards can be set up within 15 minutes.

LOVD features:


- Free, open-source, platform-independent and fully web-based tool.
- Build, curate and share LSDBs storing variants in multiple genes per patient.
- Different data access levels: website visitor, submitter, curator, database manager.
- Share submitter account with several users working at the same institute or diagnostic lab.
- Add (non-public) data fields with selection of input type (free text, numeric, etc.)
- Searching in and across data columns in all records using Boolean operators.
- Custom design of direct links (to internet, intranet or even local-PC files).
- Enhanced functionality using modules, e.g. automated mutation nomenclature check via Mutalyzer (8).
- Supports data exchange with central repositories and genome browsers (including NCBI, UCSC, EnSEMBL).

 HEALTH-200754	D4.1 User-Manual for LSDB-in-a-box V1 Software		
	WP4 - Genetics G2P databases		Security: PU
	Author(s): P. Taschner (LUMC), I. Fokkema (LUMC), C. Beroud (INSERM), J. den Dunnen (LUMC)		Version: v4.0–Final

Both UMD and LOVD were recognised as the leading systems by the participants of the Human Variome Project (5), as described in D2.1. As such, they are in a position to play an important role to solve the current lack of standardization of LSDB design and LSDB data exchange formats. To further promote the use of these ‘LSDB-in-a-box’ applications, this deliverable encompasses the user manuals for UMD (version May 2009) and LOVD (version 2.0-18)(See Annex). The manuals describe versions of the software supporting the minimal LSDB requirements listed in the ‘LSDB minimal content’ document developed by LUMC Partner and HGVS. The discussion on the ‘LSDB minimal content’ document led by WP3 is recorded on the GEN2PHEN project website and is available from <http://askja.gene.le.ac.uk/drupal5/content/lsdb-minimal-requirements>.


3. FUTURE WORK

UMD and LOVD are being developed further within WP4 to meet the needs of the community as determined by WP2 (D2.1) and in accordance with the general Gen2Phen data model developed by WP3 (D3.2). Regular updates of the ‘LSDB-in-a-box’ user manuals will be available from the UMD and LOVD websites to support the use of new functionality incorporated in new releases of the software.

 HEALTH-200754	D4.1 User-Manual for LSDB-in-a-box V1 Software		
	WP4 - Genetics G2P databases	Security: PU	
	Author(s): P. Taschner (LUMC), I. Fokkema (LUMC), C. Beroud (INSERM), J. den Dunnen (LUMC)	Version: v4.0–Final	8/9

ANNEX

The current versions of the UMD user manual (version May 2009) and the LOVD user manual (v. 2.0-18) are included in the accompanying CD to this deliverable and available on the GEN2PHEN Knowledge Centre. The latest versions of the manuals will be available from the UMD website (<http://www.umd.be/>) and the LOVD website (<http://www.lovd.nl>).

 HEALTH-200754	D4.1 User-Manual for LSDB-in-a-box V1 Software		
	WP4 - Genetics G2P databases		Security: PU
	Author(s): P. Taschner (LUMC), I. Fokkema (LUMC), C. Beroud (INSERM), J. den Dunnen (LUMC)		Version: v4.0–Final

References

1. 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.
2. 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.
3. 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.
4. Beroud C, et al. 2005. UMD (universal mutation database). *Hum Mutat* 26:184-191.
5. Cotton RG, Auerbach AD, Axton M, Barash CI, Berkovic SF, Brookes AJ, Burn J, Cutting G, den Dunnen JT, Flicek P, Freimer N, Greenblatt MS, Howard HJ, Katz M, Macrae FA, Maglott D, Möslein G, Povey S, Ramesar RS, Richards CS, Seminara D, Smith TD, Sobrido MJ, Solbakk JH, Tanzi RE, Tavtigian SV, Taylor GR, Utsunomiya J, Watson M. 2008. The Human Variome Project. *Science* 322:861-862.
6. Frederic MY, Lalande M, Boileau C, Hamroun D, Claustres M, Beroud C, Collod-Beroud G. 2009. UMD-predictor, a new prediction tool for nucleotide substitution pathogenicity-application to four genes: FBN1, FBN2, TGFBR1, and TGFBR2. *Hum Mutat* 30: 1-8.
7. Desmet FO, Hamroun D, Lalande M, Collod-Beroud G, Claustres M, Beroud C. 2009. Human Splicing Finder: an online bioinformatics tool to predict splicing signals. *Nucleic Acids Res* 37:e67.
8. Wildeman M, van Ophuizen E, van Ommen G-JB, den Dunnen JT, Taschner PE. 2008. Improving sequence variation descriptions in locus-specific mutation databases and the literature using the MUTation AnaLYZER (MUTALYZER) mutation nomenclature checker. *Hum Mutat* 29:6-13.