

Original article

hLGDB: a database of human lysosomal genes and their regulation

Alessandro Brozzi^{1,2,†}, Lorena Urbanelli^{1,†}, Pierre Luc Germain³, Alessandro Magini¹ and Carla Emiliani^{1,2,*}

¹Department of Experimental Medicine and Biochemical Sciences, University of Perugia, Via del Giochetto, 06123 Perugia, Italy,

²Centro di Eccellenza sui Materiali Innovativi Nanostrutturati (CEMIN), University of Perugia, via Elce di Sotto 8, 06123, Perugia, Italy and ³European Institute of Oncology, Via Ripamonti, Milan 20141, Italy

*Corresponding author: Tel/Fax: +00390755857436; Email: emiliani@unipg.it

†These authors contributed equally to this work.

Submitted 20 November 2012; Revised 7 February 2013; Accepted 15 March 2013

Citation details: Brozzi,A., Urbanelli,L., Germain,P.L., et al. hLGDB: a database of human lysosomal genes and their regulation. *Database* (2013) Vol. 2013: article ID bat024; doi:10.1093/database/bat024.

Lysosomes are cytoplasmic organelles present in almost all eukaryotic cells, which play a fundamental role in key aspects of cellular homeostasis such as membrane repair, autophagy, endocytosis and protein metabolism. The characterization of the genes and enzymes constituting the lysosome represents a central issue to be addressed toward a better understanding of the biology of this organelle. In humans, mutations that cause lysosomal enzyme deficiencies result in >50 different disorders and severe pathologies. So far, many experimental efforts using different methodologies have been carried out to identify lysosomal genes. The Human Lysosome Gene Database (hLGDB) is the first resource that provides a comprehensive and accessible census of the human genes belonging to the lysosomal system. This database was developed by collecting and annotating gene lists from many different sources. References to the studies that have identified each gene are provided together with cross databases gene related information. Special attention has been given to the regulation of the genes through microRNAs and the transcription factor EB. The hLGDB can be easily queried to retrieve, combine and analyze information on different lists of lysosomal genes and their regulation by microRNA (binding sites predicted by five different algorithms). The hLGDB is an open access dynamic project that will permit in the future to collapse in a unique publicly accessible resource all the available biological information about lysosome genes and their regulation.

Database URL: <http://lysosome.unipg.it/>

Introduction

Lysosomes are cellular organelles that play a pivotal role in the cell homeostasis through their involvement in degradation and recycling processes of extracellular material that has been internalized by endocytosis and intracellular components that have been sequestered by autophagy (1). Lysosomes may also fuse with the plasma membrane, emptying their contents outside the cell. This is important for processes such as cellular immune response and plasma membrane repair, both in normal and pathological

conditions (2, 3). Mutations that cause lysosomal enzyme deficiencies result in different syndromes, known as Lysosomal Storage Disorders (LSDs) (4). Most of the LSDs are associated with abnormal brain development and mental retardation. In addition, they are characterized by intracellular deposition and protein aggregation, events also found in age-related neurodegenerative disorders, such as Alzheimer's and Parkinson's diseases (5–7). These studies underline the importance of the lysosome as a central player in cell metabolism. Hence, the characterization

of genes participating in lysosomal biogenesis and function is a critical step toward the understanding of basic processes in cell biology and pathogenic mechanisms in many human diseases.

Recently, it was found that most lysosomal genes exhibit coordinated transcriptional behavior and are regulated by the transcription factor EB (TFEB), which also links autophagy to lysosome biogenesis (8, 9). Gene expression at the post-transcriptional level can be regulated by microRNAs (miRNAs). miRNAs play important roles in diverse biological processes, including development, cell differentiation, proliferation and apoptosis, in which the lysosomal system also plays an important role. Notably, miRNAs have been recently identified as involved in the regulation of autophagy (10, 11).

The Human Lysosome Gene Database (hLGDB) is the first searchable database focused on the census of genes belonging to the lysosomal system and on their regulation by miRNAs. No database resources entirely dedicated to the regulation of lysosomal genes by miRNAs or other regulators are currently available. Several lists of lysosomal genes were collected from public gene databases, published proteomics articles and reviews edited by biochemists and cell biologists working in the lysosome field. Many different algorithms are available for miRNAs binding site prediction (12–14). Five are currently present in the database. We paid special attention on balancing predictions, which were as follows: (i) more suitable to look for confirmatory evidence (TargetScanS) (15); (ii) more suitable to identify any possible target for a particular miRNA, to form

the basis for in vitro or in vivo experiments (picTar four-way and five-way) (16); (iii) more suitable to find in silico evidence for the interaction between a miRNA and a gene of a certain family or function (PITA, miRanda) (17). To increase miRNA-target mRNA information, experimentally verified miRNA targets from miRTarBase were also reported.

hLGDB aims to providing a useful resource to anyone studying the lysosomes and a tool for identifying common regulatory features of lysosomal genes. hLGDB provides a user-friendly interface through which information can be easily retrieved, including the union and intersection of different gene lists, searches for miRNA predictions and visualization on the gene transcript sequence of the miRNA target predictions.

Database Construction

The data reported in the current version 1.1, derived from NCBI PubMed searches for review articles regarding human (8, 18–25) and murine (22, 26–29) proteome of the lysosome and from lists of lysosomal genes present in the Gene Ontology (30), KEGG (31), Reactome (32) and UniProt databases (33) [Uniprot: 'Lysosome (KW-0458)' AND organism: 'Homo sapiens (Human) (9606)'; KEGG: 'Lysosome (ko04142)'; GO: GO:0005764 data stamp from the source 20120303]. The references listed within each review article (34, 35) were examined and lists of genes were extracted from either the full text or the supplementary information following a manual curation.

The screenshot displays the hLGDB interface with the following sections:

- Header:** "The Human Lysosome Gene Database" with a description: "The Lysosome Gene Database is an ongoing effort to catalogue genes encoding lysosomal proteins. For each lysosomal gene, LGDB provides information on: miRNA-genes interactions, master gene transcription factor EB (TFEB) CLEAR binding motif and binding sites." A "Human" dropdown menu is visible.
- Filter genes:**
 - View only genes reported lysosomal by **any (union)** of the following sources:
 - Proteomics Studies (view)** - 9 out of 9 selected (select all | none)
 - Databases (view)** - 4 out of 4 selected (select all | none)
 - Literature Reviews (view)** - 2 out of 2 selected (select all | none)
 - The proteome of lysosomes. Schröder BA, Proteomics 2010. Pubmed
 - Proteomics of the lysosome. Lübke T, Biochim Biophys Acta 2009. Pubmed
 - Systems Biology Approaches (view)** - 1 out of 1 selected (select all | none)
 - Filter by HGNC gene symbol:
 - TFEB binding site (max ±2kb)
 - bp upstream TSS
 - bp downstream TSS
 - CLEAR motif
 - download results in delimited format
- Filter algorithms:**
 - Predicted by **any** of the following softwares: (select all | none)
 - miRanda
 - pictar4w
 - pictar5w
 - PITA
 - TargetScan
 - miRTarBase* (*exp. verified targets)
 - Order**
 - Order results by **Gene**
- Filter miRNAs:**
 - Bound by **any** of the following miRNAs (use Ctrl+click; select all | none)
 - miRNA families:
 - let-7/98/4456/47
 - miR-10abc/10a
 - miR-124/124ab
 - miR-1255ab
 - miR-125a-5p/1
 - miR-130ac/301
 - miR-132/212/2
 - miR-1323/548a
 - miR-133abc
 - miR-135ab/135
 - miR-141/200a
 - miR-146ac/146
 - hsa-miR-10
 - hsa-miR-100
 - hsa-miR-101
 - hsa-miR-103
 - hsa-miR-105
 - 0 miRNA selected
 - [or paste a custom list](#)


Footer: hLGDB v.1.2 - © 2013 - University of Perugia, Dept. of Experimental Medicine and Biochemical Sciences | [Help](#) | [Contact](#)

Figure 1. Home page of hLGDB with search options.

The Human Lysosome Gene Database

The Lysosome Gene Database is an ongoing effort to catalogue genes encoding lysosomal proteins. For each lysosomal gene, LGDB provides information on:

- miRNA-genes interactions
- master gene transcription factor EB (TFEB) CLEAR binding motif and binding sites

Human 

 Return

[Go to mouse homolog:](#)
[Ctsf](#)

CTSF - cathepsin F

Entrez Id: [8722](#)
 HGNC Symbol: CTSF
 Name: cathepsin F
 Location: 11q13
 Other designations: CATSF, CLN13
 Transcripts: [NM_003793](#)
 Molecular Function: Catabolic enzyme
 KEGG Class: Lysosomal acid hydrolases
 KEGG Subclass: proteases
 Localization: Soluble

TFEB binding site near TSS: none within ±2kb
 CLEAR binding motif near TSS: none within ±2kb

Summary:

Cathepsins are papain family cysteine proteinases that represent a major component of the lysosomal proteolytic system. Cathepsins generally contain a signal sequence, followed by a propeptide and then a catalytically active mature region. The very long (251 amino acid residues) proregion of the cathepsin F precursor contains a C-terminal domain similar to the pro-segment of cathepsin L-like enzymes, a 50-residue flexible linker peptide, and an N-terminal domain predicted to adopt a cystatin-like fold. The cathepsin F proregion is unique within the papain family cysteine proteases in that it contains this additional N-terminal segment predicted to share structural similarities with cysteine protease inhibitors of the cystatin superfamily. This cystatin-like domain contains some of the elements known to be important for inhibitory activity. CTSF encodes a predicted protein of 484 amino acids which contains a 19 residue signal peptide. Cathepsin F contains five potential N-glycosylation sites, and it may be targeted to the endosomal/lysosomal compartment via the mannose 6-phosphate receptor pathway. The cathepsin F gene is ubiquitously expressed, and it maps to chromosome 11q13, close to the gene encoding cathepsin W. [provided by RefSeq, Jul 2008]

Uniprot function annotation:

[Q5UBX1](#) Thiol protease which is believed to participate in intracellular degradation and turnover of proteins. Has also been implicated in tumor invasion and metastasis.

Gene Ontology annotations (view)

Molecular function	Biological process	Cellular component
	proteolysis	lysosome
	antigen processing and presentation of exogenous peptide antigen via MHC class II	lysosomal lumen

Reported lysosomal by 9 sources (view)

Predicted miRNA bindings (9):

View miRNA bindings sites on the transcript sequence: [NM_003793](#)

miRNA	miRanda	pictar4w	pictar5w	PITA	TargetScan	miRTarBase*	Count
hsa-miR-137	✓			✓	✓		3
hsa-miR-140-5p	✓						1
hsa-miR-181a	✓						1
hsa-miR-18a	✓						1
hsa-miR-18b	✓						1

Figure 2. Gene information using 'CTSF' as example. (a) Summary of gene-centered information. (b) View of miRNA-binding sites on gene transcript.

hLGDB currently contains 435 genes. There are 16 sources of information divided in four main categories: Proteomic Studies, Databases, Reviews and System Biology Approaches. Each gene has been associated to its Official HGNC Gene Symbol (36) and to its Entrez Gene ID (mappings were based on data provided by Entrez with a date stamp from the source of 7 March 2012). The gene transcripts associated to each gene are annotated accordingly to NCBI RefSeq or GenBank (release 57). miRNA target predictions were extracted from the tables downloaded from the websites of the different algorithms used to predict the binding between miRNA and gene transcripts. Coordinated Lysosomal Expression and Regulation (CLEAR) is a nucleotide motif (GTACGTGAC) found to be highly enriched in the promoter set of lysosomal genes (8). We mapped this motif on both strands on the human genome (hg19) by means of fuzznuc utility of the EMBOSS package allowing one single mismatch (37). The binding sites of the TFEB come from a Chip-seq experiment carried out on HeLa cell lines (38).

hLGDB is a MySQL 5.0.95 database (constructed in the fourth normal form, some redundancy being kept to increase retrieval performance), and the interface is built in PHP.

Database Description and Utility

Search for gene lists: hLGDB can be used to retrieve and combine lists of lysosomal genes from different sources. The ANY and ALL options allow the user to either merge or intersect lists. Search for a gene using the gene symbol is also allowed (Figure 1).

Search for miRNA targets: once the user has selected or created a gene list, he/she can find miRNA (or families of miRNA) targets choosing different combinations of prediction softwares. Results are returned in a table showing information about the gene (Gene Symbol and Gene Name) and the miRNA (identifier and number of softwares predicting each binding). Gene Symbols are hyperlinked to gene-centered page where additional gene information is

```

a c t g a a g a g g g g c c c c c a g c t c g g g a c c t g g t g c t
      hsa-miR-29b
      hsa-miR-29c
      hsa-miR-29a
g a t c a g a g t g g c t g c t g c c c a g c c t g a c a t g t g t
c c a g g c c c c t c c c c g g g a g g t a c a g c t g g c a g a g g
g a a a g g c a c t g g g t a c c t c a g g g t g a g c a g a g g g c
a c t g g g c t g g g g c a c a g c c c c t g c t t c c c t g c a c c
                                     hsa-miR-18b
                                     hsa-miR-18a
c c a t t c c c a c c c t g a a g t t c t g c a c c t g c a c c t t t
hsa-miR-181a
g t t g a a t t g t g g t a g c t t a g g a g g a t g t c g g g g t g
a a g g g t g g t a t c t t g g c a g t t g a a g c t g g g g c a a g
a a c t c t g g g c t t g g g t a a t g a g c a g g a a g a a a a t t
hsa-miR-383
t t c t g a t c t t a a g c c c a g c t c t g t t c t g c c c c c g c
t t t c c t c t g t t t g a t a c t a t a a a t t t t c t g g t t c c
c t t g g a t t t a g g g a t a g t g t c c c t c t c c a t g t c e a
      hsa-miR-140-5p
                                     hsa-miR-137
                                     hsa-miR-137
                                     hsa-miR-137
g g a a a c t t g t a a c c a c c c t t t t c t a a c a g c a a t a a
a g a g g t g t c c t t g t c c c g a a a a a a a a a a a a a a a a
a a

```

Figure 2. Continued.

provided (Figure2a). On each gene transcript associated to the selected gene, miRNA binding sites are annotated (Figure2b).

Search for TFEB binding sites/CLEAR motifs: once the user has selected or created a list, he/she can choose the filter and find the lists of genes with TFEB binding sites or CLEAR motifs within a range around the transcription start site (TSS). The range is user-defined specifying downstream and upstream boundaries around the TSS. Results are provided in the table showing gene-centered information, as the hit count and distance from TSS of TFEB binding sites/CLEAR motifs.

All data in hLGDB are freely available for download as tab-delimited text files without password protection for any user. Concerning other organisms, currently the database provides a parallel orthology annotation for mouse: the user can select the species of interest using the upper right botton.

Discussion and Future Direction

hLGDB is a database that focuses on human lysosomal genes. It collects information about these genes and their transcriptional regulation such as TFEB binding sites and miRNAs. hLGDB was designed to become a lysosomal gene census. When new lysosomal genes will be discovered they will be added when the database is updated. Lysosomal genes include lysosomal hydrolases, lysosomal membrane proteins, lysosomal proteins involved in acidification and non-lysosomal proteins fundamental for this organelle biogenesis. Currently >50 recessive inherited diseases are associated with lysosomal gene dysfunction. In addition, there is increasing evidence that lysosome genes play a role in the pathogenesis of common neurodegenerative diseases such as Alzheimer's, Parkinson's and Huntington's.

Researchers may benefit from hLGDB because they have in a single reference to the broadest compendium of

lysosomal gene lists. They can search for miRNA targets combining up to six different methods. Results of miRNA targets may be directly compared with other transcriptional regulation elements such as the distance from the TSS of TFEB binding site or the distance to a CLEAR sequence to identify common features of regulation.

hLGDB has been designed to integrate additional layers of biological information, such as experimental data and comparative genomics. Currently the database present information for human and mouse species; in the next versions, additional species, such as rat, will be integrated. Finally, hLGDB provides a powerful resource to system biology approaches and network analysis to dissect the map of interactions taking place in the lysosomal system.

Funding

This work was supported by ELA (European Leukodystrophies Association) Grant no. 2011-037C1B and Fondazione Cassa di Risparmio di Perugia Grant no. 2010.011.0434. Funding for open access charge: Fondazione Cassa di Risparmio di Perugia Grant no. 2010.011.0434.

Conflict of interest. None declared.

References

- Kornfeld,S. and Mellman,I. (1989) The biogenesis of lysosomes. *Annu. Rev. Cell Biol.*, **5**, 483–525.
- Reddy,A., Caler,E.V. and Andrews,N.W. (2001) Plasma membrane repair is mediated by Ca(2+)-regulated exocytosis of lysosomes. *Cell*, **106**, 157–169.
- Stinchcombe,J., Bossi,G. and Griffiths,G.M. (2004) Linking albinism and immunity: the secrets of secretory lysosomes. *Science*, **305**, 55–59.
- Ballabio,A. and Gieselmann,V. (2009) Lysosomal disorders: from storage to cellular damage. *Biochim. Biophys. Acta*, **1793**, 684–696.
- Urbanelli,L., Magini,A., Polchi,A. et al. (2011) Recent developments in therapeutic approaches for lysosomal storage diseases. *Recent Pat. CNS Drug. Discov.*, **6**, 1–19.
- Bahr,B.A. and Bendiske,J. (2002) The neuropathogenic contributions of lysosomal dysfunction. *J. Neurochem.*, **83**, 481–489.
- Emiliani,C., Urbanelli,C., Racanicchi,L. et al. (2003) Up-regulation of glycohydrolases in Alzheimer's disease fibroblasts correlates with Ras activation. *J. Biol. Chem.*, **278**, 38453–38460.
- Sardiello,M., Palmieri,M., di Ronza,A. et al. (2009) A gene network regulating lysosomal biogenesis and function. *Science*, **325**, 473–477.
- Settembre,C., Di Malta,C., Polito,V.A. et al. TFEB links autophagy to lysosomal biogenesis. *Science*, **332**, 1429–1433.
- Frankel,L.B., Wen,J., Lees,M. et al. microRNA-101 is a potent inhibitor of autophagy. *EMBO J.*, **30**, 4628–4641.
- Zhu,H., Wu,H., Liu,X. et al. (2009) Regulation of autophagy by a beclin 1-targeted microRNA, miR-30a, in cancer cells. *Autophagy*, **5**, 816–823.
- Shirdel,E.A., Xie,W., Mak,T.W. et al. NAViGaTing the micronome—using multiple microRNA prediction databases to identify signalling pathway-associated microRNAs. *PLoS One*, **6**, e17429.
- Sethupathy,P., Corda,B. and Hatzigeorgiou,A.G. (2006) TarBase: A comprehensive database of experimentally supported animal microRNA targets. *RNA*, **12**, 192–197.
- Maziere,P. and Enright,A.J. (2007) Prediction of microRNA targets. *Drug. Discov. Today*, **12**, 452–258.
- Lewis,B.P., Burge,C.B. and Bartel,D.P. (2005) Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell*, **120**, 15–20.
- Grun,D., Wang,Y.L., Langenberger,D. et al. (2005) microRNA target predictions across seven Drosophila species and comparison to mammalian targets. *PLoS Comput. Biol.*, **1**, e13.
- Betel,D., Wilson,M., Gabow,A. et al. (2008) The microRNA.org resource: targets and expression. *Nucleic Acids Res.*, **36**, D149–D53.
- Chataway,T.K., Whittle,A.M., Lewis,M.D. et al. (1998) Two-dimensional mapping and microsequencing of lysosomal proteins from human placenta. *Placenta*, **19**, 643–654.
- Journet,A., Chapel,A., Kieffer,S. et al. (2000) Towards a human repertoire of monocytic lysosomal proteins. *Electrophoresis*, **21**, 3411–3419.
- Journet,A., Chapel,A., Kieffer,S. et al. (2002) Proteomic analysis of human lysosomes: application to monocytic and breast cancer cells. *Proteomics*, **2**, 1026–1040.
- Sleat,D.E., Lackland,H., Wang,Y. et al. (2005) The human brain mannose 6-phosphate glycoproteome: a complex mixture composed of multiple isoforms of many soluble lysosomal proteins. *Proteomics*, **5**, 1520–1532.
- Sleat,D.E., Zheng,H., Qian,M. et al. (2006) Identification of sites of mannose 6-phosphorylation on lysosomal proteins. *Mol. Cell Proteomics*, **5**, 686–701.
- Sleat,D.E., Wang,Y., Sohar,I. et al. (2006) Identification and validation of mannose 6-phosphate glycoproteins in human plasma reveal a wide range of lysosomal and non-lysosomal proteins. *Mol. Cell Proteomics*, **5**, 1942–1956.
- Schroder,B., Wrocklage,C., Pan,C. et al. (2007) Integral and associated lysosomal membrane proteins. *Traffic*, **8**, 1676–1686.
- Sleat,D.E., Zheng,H. and Lobel,P. (2007) The human urine mannose 6-phosphate glycoproteome. *Biochim. Biophys. Acta*, **1774**, 368–3672.
- Kollmann,K., Mutenda,K.E., Balleininger,M. et al. (2005) Identification of novel lysosomal matrix proteins by proteome analysis. *Proteomics*, **5**, 3966–3978.
- Czapalla,C., Mansukoski,H., Riedl,T. et al. (2006) Proteomic analysis of lysosomal acid hydrolases secreted by osteoclasts: implications for lytic enzyme transport and bone metabolism. *Mol. Cell Proteomics*, **5**, 134–143.
- Zhang,H., Fan,X., Bagshaw,R.D. et al. (2007) Lysosomal membranes from beige mice contain higher than normal levels of endoplasmic reticulum proteins. *J. Proteome Res.*, **6**, 240–249.
- Qian,M., Sleat,D.E., Zheng,H. et al. (2008) Proteomics analysis of serum from mutant mice reveals lysosomal proteins selectively transported by each of the two mannose 6-phosphate receptors. *Mol. Cell Proteomics*, **7**, 58–70.
- Ashburner,M., Ball,C.A., Blake,J.A. et al. (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.*, **25**, 25–29.
- Kanehisa,M. and Goto,S. (2000) KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.*, **28**, 27–30.

-
32. Vastrik,I., D'Eustachio,P., Schmidt,E. *et al.* (2007) Reactome: a knowledge base of biologic pathways and processes. *Genome Biol.*, **8**, R39.
33. Leinonen,R., Diez,F.G., Binns,D. *et al.* (2004) UniProt archive. *Bioinformatics*, **20**, 3236–3237.
34. Schroder,B.A., Wrocklage,C., Hasilik,A. *et al.* The proteome of lysosomes. *Proteomics*, **10**, 4053–4076.
35. Lubke,T., Lobel,P. and Sleat,D.E. (2009) Proteomics of the lysosome. *Biochim. Biophys. Acta*, **1793**, 625–635.
36. Seal,R.L., Gordon,S.M., Lush,M.J. *et al.* genenames.org: the HGNC resources in 2011. *Nucleic Acids Res.*, **39**, D514–D519.
37. Rice,P., Longden,I. and Bleasby,A. (2000) EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet.*, **16**, 276–277.
38. Palmieri,M., Impey,S., Kang,H. *et al.* Characterization of the CLEAR network reveals an integrated control of cellular clearance pathways. *Hum. Mol. Genet.*, **20**, 3852–3866.
-