hLGDB: a database of human lysosomal genes and their regulation

Alessandro Brozzi1,2,*, Lorena Urbanelli1,1, Pierre Luc Germain3, Alessandro Magini1 and Carla Emiliani1,2,∗

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, endocitosis 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 identity 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

© The Author(s) 2013. Published by Oxford University Press. This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/3.0/), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.
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.

Figure 1. Home page of hLGDB with search options.
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 prediction softwares. Coordinated Lysosomal Expression and Regulation (CLEAR) is a nucleotide motif (GTCACGTGAC) 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

---

**Figure 2.** Gene information using ‘CTSF’ as example. (a) Summary of gene-centered information. (b) View of miRNA-binding sites on gene transcript.
provided (Figure 2a). On each gene transcript associated to the selected gene, miRNA binding sites are annotated (Figure 2b).

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 button.

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
lyosomal 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.

hLGBDB has been designed to integrate additional layers of biological information, such as experimental data and comparative genomics. Currently the database presents information for human and mouse species; in the next versions, additional species, such as rat, will be integrated. Finally, hLGBD 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
1. Kornfeld,S. and Mellman,I. (1989) The biogenesis of lysosomes. Annu. Rev. Cell Biol., 5, 483–525.
2. 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.
3. Stinchcombe,J., Bossi,G. and Griffiths,G.M. (2004) Linking albinism and immunity: the secrets of secretory lysosomes. Science, 305, 55–59.
4. Ballabio,A. and Gieselmann,V. (2009) Lysosomal disorders: from storage to cellular damage. Biochem. Biophys. Acta, 1793, 684–696.
5. 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.
6. Bahr,B.A. and Bendiske,J. (2002) The neuropathogenic contributions of lysosomal dysfunction. J. Neurochem., 83, 481–489.
7. Emiliian,C., Urbanelli,C., Racinicli,L. et al. (2003) Up-regulation of glycohydrolases in Alzheimer’s disease fibroblasts correlates with Ras activation. J. Biol. Chem., 278, 38453–38460.
8. Sardiello,M., Palmieri,M., di Ronza,A. et al. (2004) Linking albinism and immunity: the secrets of secretory lysosomes. Science, 305, 55–59.
9. Settembre,C., Di Maite,C., Polito,V.A. et al. TFEB links autophagy to lysosomal biogenesis. Function, Science, 325, 473–477.
10. Franke,L.B., Wen,J., Lees,M. et al. microRNA-101 is a potent inhibitor of autophagy. EMBO J., 30, 4626–4641.
11. 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.
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.