Database, Vol. 2011, Article ID bar055, doi:10.1093/database/bar055

Original article

hUbiquitome: a database of experimentally verified ubiquitination cascades in humans

Yipeng Du[1,2,†], Nanfang Xu[1,3,†], Ming Lu[1] and Tingting Li[1,4,*]

[1]Department of Medical Informatics, [2]Key Laboratory of Carcinogenesis and Translational Research (Ministry of Education), Department of Biochemistry and Molecular Biology, Peking University Health Science Center, Beijing 100191, [3]Peking University First Hospital, Beijing 100034 and [4]Institute of Systems Biomedicine, Peking University Health Science Center, Beijing 100191, China

***Corresponding author:** Tel: +86 10 8280 1585; Fax: +86 10 8280 1001; Email: litt@hsc.pku.edu.cn

[†]These authors contributed equally to this work.

Submitted 25 August 2011; Revised 27 October 2011; Accepted 8 November 2011

Protein ubiquitination is an evolutionarily conserved and functionally diverse post-translational modification achieved through the sequential action of E1-activating enzymes, E2-conjugating enzymes and E3 ligases. A summary of validated ubiquitination substrates have been presented and a prediction of new substrates have been conducted in yeast. However, a systematic summary of human ubiquitination substrates containing experimental evidence and the enzymatic cascade of each substrate is not available. In the present study, hUbiquitome web resource is introduced, a public resource for the retrieval of experimentally verified human ubiquitination enzymes and substrates. hUbiquitome is the first comprehensive database of human ubiquitination cascades. Currently, hUbiquitome has in its repertoire curated data comprising 1 E1 enzyme, 12 E2 enzymes, 138 E3 ligases or complexes, 279 different substrate proteins and 17 deubiquitination enzyme terms. The biological functions of substrates from different kinds of E3s were analyzed using the collected data. The findings show that substrates ubiquitinated by RING (Really Interesting New Gene) E3s are enriched most in apoptosis-related processes, whereas substrates ubiquitinated by other E3s are enriched in gene expression-associated processes. An analysis of the data demonstrates the biological process preferences of the different kinds of E3s. hUbiquitome is the first database to systematically collect experimentally validated ubiquitinated proteins and related ubiquitination cascade enzymes which might be helpful in the field of ubiquitination-modification research.
Database URL: http://202.38.126.151/hmdd/hubi/

Introduction

Ubiquitination is an important type of post-translational modification in which an isopeptide bond is formed between the C-terminus of ubiquitin and a lysine residue from either a substrate or another ubiquitin molecule (1). In addition to its original role in protein degradation (2), ubiquitination regulates other cellular processes, including transcription, cell cycle, DNA repair, apoptosis and receptor endocytosis (3). Thus, aberrations of ubiquitin–proteasome system function in all the above-mentioned processes have been implicated in the pathogenesis of human diseases, ranging from inflammatory, neurodegenerative muscle-wasting, to various forms of malignancies (4, 5).

A few resources of ubiquitination data are available. Ubiprot is the first database on ubiquitination (6), which focuses on the properties of ubiquitinated proteins *per se*, with data from different species combined. The data in Ubiprot are mainly sourced from several high-throughput studies, and do not include information of ubiquitination cascade enzymes. The other ubiquitination database is a yeast database [Saccharomyces cerevisiae Ubiquitination Database (SCUD)] focusing on ubiquitination cascades (7). SCUD has collected almost all known enzymes involved in the ubiquitination process of yeast and has grouped them into reasonable classes. E3Miner is a text-mining tool for literature search (8), and appears to replace labor-intense manual curation with machine learning approach.

Database, Vol. 2011, Article ID bar055, doi:10.1093/database/bar055

Although E3Miner contains many useful E3-centered information, obtaining high accuracy in the text-mining method is difficult. Ubiquitination is a conserved biological process from yeasts to humans (1, 9–12). Understanding of the human ubiquitination system has gradually become clearer as more enzymes and ubiquitinated proteins have been discovered (13–15). However, no database demonstrating those of the human ubiquitination system exists.

hUbiquitome is the first and largest searchable collection of human ubiquitination proteins and cascades. hUbiquitome was constructed based on published papers in PubMed, and all the ubiquitination proteins and cascades were experimentally validated. hUbiquitome provides a user-friendly interface through which information can be easily retrieved, including E1, E2 and E3 substrates, deubiquitination enzymes (DUBs) and the relationship among these elements. Ubiquitin lysine sites or sequences are provided if they were identified in the reference papers. If some E3 functions are in the form of complexes, a complex name would be provided. hUbiquitome aims to provide as many precise information as confirmed by experiments in the ubiquitination cascades.

Database construction

The ubiquitination cascade data documented in the current version, hUbiquitome 1.1.1, were collected manually by searching the PubMed database for primary research articles published or e-published up to the time of the present study with a list of keywords (e1 ubiquitin OR e2 ubiquitin-conjugating enzyme OR e3 ubiquitin ligases Limits: Humans). Full articles and supplementary data were examined. When appropriate, references were checked for additional materials. Studies that described experimentally identified relationships between E2 and E3 and E3 and Substrate were included. Only papers showing adequate experimental evidence of substrate ubiquitination by identified E3s were selected. Experimental evidence includes *in vitro* reaction of E3 and substrate, immunoblot with ubiquitin antibody, autoradiogram of isotope-labeled ubiquitin, tag-labeled ubiquitin detection, substrate degradation by proteome, substrate stability detection and so on. Additional information, such as E3 complex components and ubiquitinated sites and sequences, are important features of hUbiquitome.

Another part of the ubiquitination system is the deubiquitination process (16–19). DUBs are involved in multiple cellular processes similar to ubiquitinating enzymes (UBs). However, studies on DUBs lag behind that of UBs. Seventeen terms of DUBs with identified substrates were collected and added after E3-substrate cascade terms.

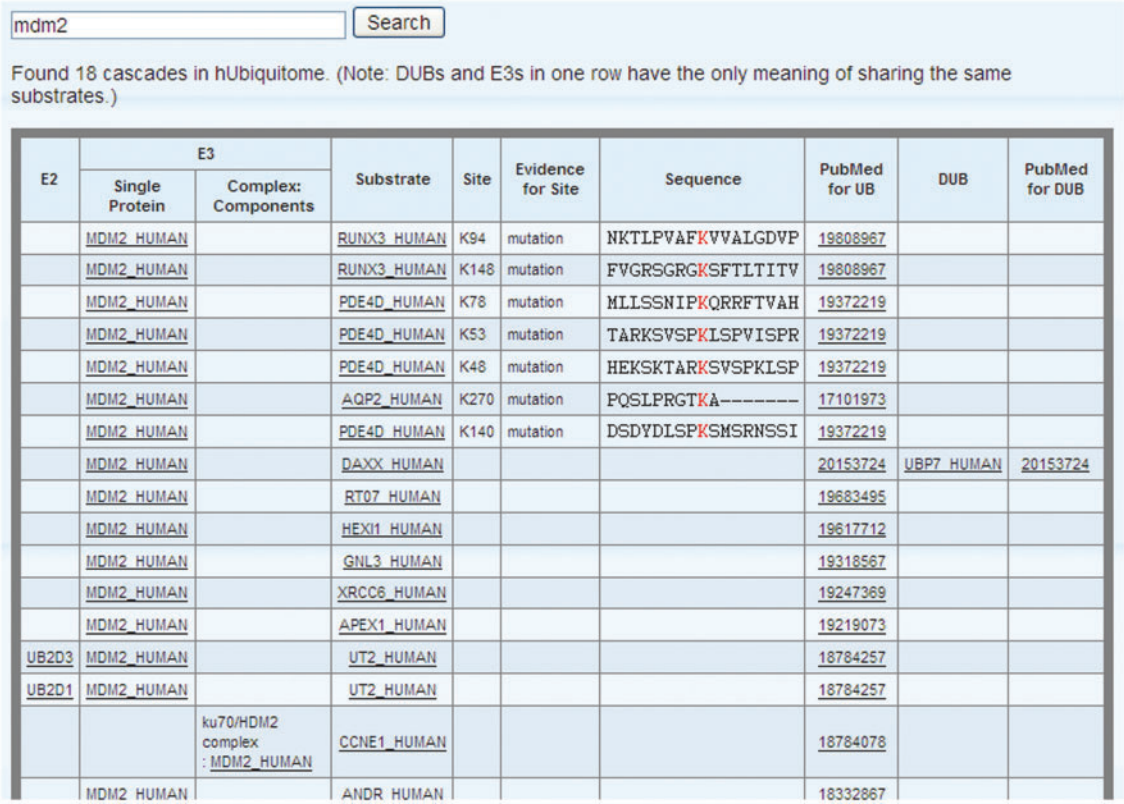
hUbiquitome 1.1.1 currently contains 1 E1 enzyme, 12 E2 enzymes, 138 E3 ligases or E3 ligases complexes, 279 different substrate proteins and 17 DUB terms. All the proteins and the related papers in the website are hyperlinked to UniProt and PubMed.

Database description and utility

Search: hUbiquitome can be used to search for an enzyme or substrate using the UniprotKB protein accession number or the UniprotKB protein entry name. For example, input 'Mdm2' inside the searchbox and search. A list of cascaded contexts for Mdm2 appears (Figure 1). E1 is common to all the cascades, so it is omitted. Papers reporting E3-substrates reactions do not always inculde E2 information. To be precise, only E2 information in the cascades reported in previous papers are shown. However, users can deduce that the same E2 could be used with other E3-substrates reactions that share the same E3. For example, a paper (PMID:18784257) reported that Mdm2 ubiquitinate UT2 uses UB2D1 as E2. Thus, UB2D1 might also be used as E2 in Mdm2-mediated ubiquitination of other substrates, such as RUNX3, PDE4D and AQP2. In some cases, E3 functions in the form of a complex. For example, ku70 and Mdm2 function together to ubiquitinate CCNE1. A complex name is given if E3 functions as a complex. Some experiments identified the ubiquitinated lysine sites and sequence motifs, such as Mdm2 ubiquitinate RUNX3 at lysines 94 and 148. The experimentally identified ubiquitinated lysine sites and sequence motifs are given as soon as they are available. The experimental evidence (mass spectrometry, mutation or both) identifying the ubiquitinated lysine sites are also included. Although the number of DUBs is small than that of E3s, we include them in the cascades, for example, UBP7 have been reported to be a DUB for DAXX in a paper (PMID: 20153724). MDM2 and UBP7 arranged in one row has the only meaning of sharing the same substrates. The PubMed ID is hyperlinked to its source page at NCBI, and each protein is ultimately hyperlinked to major biological databases like Uniprot and Entrez Gene.

Blast: the hUbiquitome Blast reports lysine sites in submitted sequences that match with ubiquitinated lysine sites in the hUbiquitome database. Thirty-five unique lysine sites were found to be ubiquitinated by known E3s (excluding those identified by mass spectrometry without E3 information). The hUbiquitome Blast is useful for retrieving possible ubiquitinated lysine sites of interesting proteins. However, Blast does not give meaningful significant values. Thus, the biological context of the Blast results should be evaluated by the user.

Submit and Download: Users can submit their own proteins or even entire ubiquination cascades to hUbiquitome. Contributors just need to fill out the Excel form provided at the Submission Page and then send it back to the database administrator.



mdm2 Search

Found 18 cascades in hUbiquitome. (Note: DUBs and E3s in one row have the only meaning of sharing the same substrates.)

E2	E3		Substrate	Site	Evidence for Site	Sequence	PubMed for UB	DUB	PubMed for DUB
	Single Protein	Complex: Components							
	MDM2_HUMAN		RUNX3_HUMAN	K94	mutation	NKTLPVAFKVVALGDVP	19808967		
	MDM2_HUMAN		RUNX3_HUMAN	K148	mutation	FVGRSGRGKSFTLTITV	19808967		
	MDM2_HUMAN		PDE4D_HUMAN	K78	mutation	MLLSSNIPKQRRFTVAH	19372219		
	MDM2_HUMAN		PDE4D_HUMAN	K53	mutation	TARKSVSPKLSPVISPR	19372219		
	MDM2_HUMAN		PDE4D_HUMAN	K48	mutation	HEKSKTARKSVSPKLSP	19372219		
	MDM2_HUMAN		AQP2_HUMAN	K270	mutation	PQSLPRGTKA-------	17101973		
	MDM2_HUMAN		PDE4D_HUMAN	K140	mutation	DSDYDLSPKSMSRNSSI	19372219		
	MDM2_HUMAN		DAXX_HUMAN				20153724	UBP7_HUMAN	20153724
	MDM2_HUMAN		RT07_HUMAN				19683495		
	MDM2_HUMAN		HEXI1_HUMAN				19617712		
	MDM2_HUMAN		GNL3_HUMAN				19318567		
	MDM2_HUMAN		XRCC6_HUMAN				19247369		
	MDM2_HUMAN		APEX1_HUMAN				19219073		
UB2D3	MDM2_HUMAN		UT2_HUMAN				18784257		
UB2D1	MDM2_HUMAN		UT2_HUMAN				18784257		
		ku70/HDM2 complex : MDM2_HUMAN	CCNE1_HUMAN				18784078		
	MDM2_HUMAN		ANDR_HUMAN				18332867		

Figure 1. General view using 'Mdm2' as an example.

All data in hUbiquitome are freely available for download as tab-delimited text files without password protection for academic users.

Functional analysis of E3 substrates: E3 can be divided into two main classes (RING: Really Interesting New Gene and HECT: homologous to *E6-AP* C-Terminus E3) according to the domain structure they contain (13, 15). Other kinds of E3s contain few numbers, such as U-box and PHD E3s (20, 21). Whether different E3 classes prefer to catalyze functionally different substrates has never been reported. In this article, the biological functions of substrates catalyzed by RING or HECT and other E3s are analyzed using the online DAVID tool (22). RING substrates (RING) or HECT and other substrates (HECT) are submitted to the DAVID website with the background of the whole human proteome. The functional annotation tool, gene ontology biological process (GOTERM_BP), was selected to analyze the functional enrichment of the two kinds of substrates under default parameters. The enrichment was defined by Benjamini-adjusted *P*-value designed to control false discovery rates. From the enriched terms in both the RING and HECT group results, nine terms were selected according to the different enrichment and biological meanings in the two groups. Detailed information can be found in supplement materials. The findings show that apoptosis-associated processes are enriched in RING E3s (Figure 2A), whereas gene transcription regulation-associated processes are enriched in HECT and in other E3s (Figure 2B). This result indicates that different classes of E3s may prefer different biological processes.

Database implementation: hUbiquitome consists of a relational Sqlite (http://www.sqlite.org/) database and a JqueryUI (http://jquery.com/web interface), constructed in Python (http://www.python.org/) with Django (http://www.djangoproject.com/) and run via an Apache server (http://www.apache.org/).

Discussion and future direction

hUbiquitome, a database that focuses on the human ubiquitination system, collects exact ubiquitination information which have been experimentally validated. For every ubiquitinated substrate, hUbiquitome provides the other ubiquitination cascades information in one row which includes E2, E3, ubiquitinated sites and sequences if possible. Researchers can search for individual terms and blast peptide sequences. They can also download the full datasheet. Based on the collection, the biological process preferences of RING E3s and other E3s are found,

Database, Vol. 2011, Article ID bar055, doi:10.1093/database/bar055

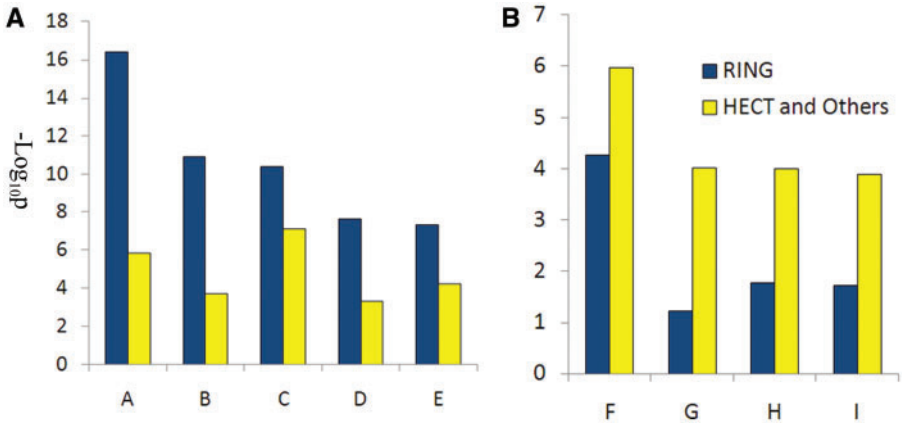


Figure 2. Functional analysis of RING E3s, HECT and other E3s. (**A**) Biological processes enriched by RING E3s; (**B**) biological processed enriched by HECT and other E3s. A, regulation of apoptosis; B, apoptosis; C, cellular response to stress; D, cell cycle; E: response to DNA damage stimulus; F: positive regulation of macromolecule metabolic process; G: positive regulation of transcription; H: positive regulation of RNA metabolic process; I: positive regulation of gene expression.

indicating biological differences among various kinds of E3s.

hUbiquitome was designed to cover all experimentally validated ubiquitination associated proteins (enzymes and substrates) and cascades in humans. More ubiquitinated proteins will soon be discovered, although the current version of hUbiquitome includes hundreds of them. These newly demonstrated ubiquitinated proteins will be added to hUbiquitome when the database is updated. hUbiquitome does not collect ubiquitinated proteins without E3 information produced by large-scale mass spectrometry methods (23–25).

Researchers may benefit from hUbiquitome in three ways. First, scholars can search for interesting proteins to find their ubiquitination cascades. Some of these proteins are even hyperlinked to the original papers to provide further information. Second, based on the characteristics of ubiquitinated peptide sequences, the BLAST function provides researchers a tool to estimate possible ubiquitinations of unknown peptides. Third, using the entire data which can be freely downloaded from the website, researchers can conduct further analysis. For example, the functional preference of two classes of E3s can be analyzed. Other interesting analyses, such as E2 and E3 substrate network construction, are expected.

Supplementary Data

Supplementary data are available at *Database* online.

Funding

National Basic Research Program of China (2011CBA01104); National Natural Science Foundation of China (60905014 and 31030041). The funding agencies played no role in the study design, data collection, analysis, decision to publish or preparation of the manuscript. Funding for open access charge: National Natural Science Foundation of China (60905014).

Conflict of interest. None declared.

References

1. Pickart,C.M. (2001) Mechanisms underlying ubiquitination. *Annu. Rev. Biochem.*, **70**, 503–533.
2. Ciechanover,A. (2005) Proteolysis: from the lysosome to ubiquitin and the proteasome. *Nat. Rev. Mol. Cell Biol.*, **6**, 79–87.
3. Reinstein,E. and Ciechanover,A. (2006) Narrative review: protein degradation and human diseases: the ubiquitin connection. *Ann. Intern. Med.*, **145**, 676–684.
4. Ciechanover,A., Orian,A. and Schwartz,A.L. (2000) The ubiquitin-mediated proteolytic pathway: mode of action and clinical implications. *J. Cell Biochem. Suppl.*, **34**, 40–51.
5. Schwartz,A.L. and Ciechanover,A. (2009) Targeting proteins for destruction by the ubiquitin system: implications for human pathobiology. *Annu. Rev. Pharmacol. Toxicol.*, **49**, 73–96.
6. Chernorudskiy,A.L., Garcia,A., Eremin,E.V. *et al*. (2007) UbiProt: a database of ubiquitylated proteins. *BMC Bioinformatics*, **8**, 126.
7. Lee,W.C., Lee,M., Jung,J.W. *et al*. (2008) SCUD: Saccharomyces cerevisiae ubiquitination database. *BMC Genomics*, **9**, 440.
8. Lee,H., Yi,G.S. and Park,J.C. (2008) E3Miner: a text mining tool for ubiquitin-protein ligases. *Nucleic Acids Res.*, **36**, W416–W422.
9. Hershko,A. and Ciechanover,A. (1998) The ubiquitin system. *Annu. Rev. Biochem.*, **67**, 425–479.
10. Hochstrasser,M., Johnson,P.R., Arendt,C.S. *et al*. (1999) The Saccharomyces cerevisiae ubiquitin-proteasome system. *Philos. Trans. R Soc. Lond. B Biol. Sci.*, **354**, 1513–1522.
11. Pickart,C.M. and Eddins,M.J. (2004) Ubiquitin: structures, functions, mechanisms. *Biochim. et Biophys. Acta Mol. Cell Res.*, **1695**, 55–72.

12. Sorokin,A.V., Kim,E.R. and Ovchinnikov,L.P. (2009) Proteasome system of protein degradation and processing. *Biochemistry*, **74**, 1411–1442.
13. Bernassola,F., Karin,M., Ciechanover,A. *et al*. (2008) The HECT family of E3 ubiquitin ligases: multiple players in cancer development. *Cancer Cell*, **14**, 10–21.
14. Cardozo,T. and Pagano,M. (2004) The SCF ubiquitin ligase: insights into a molecular machine. *Nat. Rev. Mol. Cell Biol.*, **5**, 739–751.
15. Deshaies,R.J. and Joazeiro,C.A. (2009) RING domain E3 ubiquitin ligases. *Annu. Rev. Biochem.*, **78**, 399–434.
16. Hussain,S., Zhang,Y. and Galardy,P.J. (2009) DUBs and cancer: the role of deubiquitinating enzymes as oncogenes, non-oncogenes and tumor suppressors. *Cell Cycle*, **8**, 1688–1697.
17. Nijman,S.M., Luna-Vargas,M.P., Velds,A. *et al*. (2005) A genomic and functional inventory of deubiquitinating enzymes. *Cell*, **123**, 773–786.
18. Amerik,A.Y. and Hochstrasser,M. (2004) Mechanism and function of deubiquitinating enzymes. *Biochim. Biophys. Acta*, **1695**, 189–207.
19. Katz,E.J., Isasa,M. and Crosas,B. (2010) A new map to understand deubiquitination. *Biochem. Soc. Trans.*, **38**, 21–28.
20. Coscoy,L. and Ganem,D. (2003) PHD domains and E3 ubiquitin ligases: viruses make the connection. *Trends Cell Biol.*, **13**, 7–12.
21. Hatakeyama,S. and Nakayama,K.I. (2003) U-box proteins as a new family of ubiquitin ligases. *Biochem. Biophys. Res. Commun.*, **302**, 635–645.
22. Huang da,W., Sherman,B.T. and Lempicki,R.A. (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.*, **4**, 44–57.
23. Parker,C.E., Warren,M.R., Mocanu,V. *et al*. (2008) Mass spectrometric determination of protein ubiquitination. *Methods Mol. Biol.*, **446**, 109–130.
24. Meierhofer,D., Wang,X., Huang,L. *et al*. (2008) Quantitative analysis of global ubiquitination in HeLa cells by mass spectrometry. *J. Proteome Res.*, **7**, 4566–4576.
25. Jeram,S.M., Srikumar,T., Pedrioli,P.G. *et al*. (2009) Using mass spectrometry to identify ubiquitin and ubiquitin-like protein conjugation sites. *Proteomics*, **9**, 922–934.