



Autophagy



ISSN: (Print) (Online) Journal homepage: <https://www.tandfonline.com/loi/kaup20>

LIRcentral: a manually curated online database of experimentally validated functional LIR motifs

Agathangelos Chatzichristofi, Vasileios Sagris, Aristos Pallaris, Marios Eftychiou, Ioanna Kalvari, Nicholas Price, Theodosios Theodosiou, Ioannis Iliopoulos, Ioannis P. Nezis & Vasilis J Promponas

To cite this article: Agathangelos Chatzichristofi, Vasileios Sagris, Aristos Pallaris, Marios Eftychiou, Ioanna Kalvari, Nicholas Price, Theodosios Theodosiou, Ioannis Iliopoulos, Ioannis P. Nezis & Vasilis J Promponas (2023): LIRcentral: a manually curated online database of experimentally validated functional LIR motifs, *Autophagy*, DOI: [10.1080/15548627.2023.2235851](https://doi.org/10.1080/15548627.2023.2235851)

To link to this article: <https://doi.org/10.1080/15548627.2023.2235851>



© 2023 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



[View supplementary material](#)



Published online: 02 Aug 2023.



[Submit your article to this journal](#)



Article views: 401













[View related articles](#)



[View Crossmark data](#)

LIRcentral: a manually curated online database of experimentally validated functional LIR motifs

Agathangelos Chatzichristofi ^{a,b}, Vasileios Sagris ^b, Aristos Pallaris ^b, Marios Eftychiou ^{b,#}, Ioanna Kalvari ^b, Nicholas Price ^b, Theodosios Theodosiou ^b, Ioannis Iliopoulos ^a, Ioannis P. Nezis ^c, and Vasilis J Promponas ^b

^aDivision of Basic Sciences, School of Medicine, University of Crete, Heraklion, Crete, Greece; ^bBioinformatics Research Laboratory, Department of Biological Sciences, University of Cyprus, Nicosia, Cyprus; ^cSchool of Life Sciences, University of Warwick, Coventry, UK

ABSTRACT

Several selective macroautophagy receptor and adaptor proteins bind members of the Atg8 (autophagy related 8) family using short linear motifs (SLiMs), most often referred to as Atg8-family interacting motifs (AIMs) or LC3-interacting regions (LIRs). AIM/LIR motifs have been extensively studied during the last fifteen years, since they can uncover the underlying biological mechanisms and possible substrates for this key catabolic process of eukaryotic cells. Prompted by the fact that experimental information regarding LIR motifs can be found scattered across heterogeneous literature resources, we have developed LIRcentral (<https://lircentral.eu>), a freely available online repository for user-friendly access to comprehensive, high-quality information regarding LIR motifs from manually curated publications. Herein, we describe the development of LIRcentral and showcase currently available data and features, along with our plans for the expansion of this resource. Information incorporated in LIRcentral is useful for accomplishing a variety of research tasks, including: (i) guiding wet biology researchers for the characterization of novel instances of LIR motifs, (ii) giving bioinformaticians/computational biologists access to high-quality LIR motifs for building novel prediction methods for LIR motifs and LIR containing proteins (LIRCPs) and (iii) performing analyses to better understand the biological importance/features of functional LIR motifs. We welcome feedback on the LIRcentral content and functionality by all interested researchers and anticipate this work to spearhead a community effort for sustaining this resource which will further promote progress in studying LIR motifs/LIRCPs.

Abbreviations: AIM, Atg8-family interacting motif; Atg8, autophagy related 8; GABARAP, GABA type A receptor-associated protein; LIR, LC3-interacting region; LIRCP, LIR-containing protein; MAP1LC3/LC3, microtubule associated protein 1 light chain 3; PMID, PubMed identifier; PPI, protein-protein interaction; SLiM, short linear motif

ARTICLE HISTORY

Received 16 June 2022
Revised 14 June 2023
Accepted 6 July 2023

KEYWORDS

Atg8 family; database; manual literature curation/annotation; online resource; peptide-protein interaction; selective macroautophagy



Introduction


Autophagy comprises lysosome-mediated degradation processes for recycling cytoplasmic material and plays a central role in eukaryotic cell homeostasis and stress response [1]. Macroautophagy is the best studied form of autophagy; it relies on the nucleation, and expansion and maturation of phagophores – transient double-membraned structures engulfing cytoplasmic material to be degraded – which become autophagosomes that eventually fuse with lysosomes for recycling to occur [2]. Key components of the macroautophagy machinery are members of the Atg8 protein family (MAP1LC3/LC3 and GABARAP in human), small proteins with a ubiquitin-like fold: they participate essentially in all steps of macroautophagy from phagophore nucleation, to autolysosome formation [3]. These proteins are involved in an increasing number of biological processes, including LC3-associated phagocytosis [4,5], endosomal microautophagy [5,6], and unconventional secretion [7].

In selective macroautophagy, selectivity is mainly mediated through receptor proteins that bind to the surface of Atg8s by

anchoring a small linear peptide (namely the Atg8-family interacting motif [AIM] in plants and fungi and the LC3-interacting region [LIR] in mammals [8]). Following the discovery of the first few LIR-containing proteins (LIRCPs) [9–11], several definitions of the LIR motif have appeared in the literature [9,12–15], highlighting a short core consensus sequence described by the regular expression pattern [WFY]xx[VLI]. Similar to other cases of short linear motifs (SLiMs), LIR motifs are often located within intrinsically disordered proteins [16–18]; this notion has already been exploited in reducing false positive hits in LIR-motif detection tools [19].

The collection of a few dozen experimentally characterized LIR motifs made it possible to develop computational methods for their prediction [19,20]. Such prediction tools have inevitably facilitated the advancement of research toward identifying the molecular mechanisms underlying several subtypes of selective macroautophagy and related biological processes [21], and the number of scientific publications reporting the experimental validation of functional LIR motifs

CONTACT Vasilis J Promponas  vprobon@ucy.ac.cy  Bioinformatics Research Laboratory, Department of Biological Sciences, University of Cyprus, Nicosia, Cyprus
[#]Present address: Leuven Statistics Research Centre (LStat), KU Leuven, Belgium.

 Supplemental data for this article can be accessed online at <https://doi.org/10.1080/15548627.2023.2235851>

© 2023 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

continues to accumulate. Despite the importance of experimentally verified LIR motifs, for wet lab biologists and computational biologists alike, there is a lack of a central repository of LIR-motif instances curated from the literature describing such sequences across species.

Currently, researchers wishing to obtain information about LIR motifs have access to online prediction tools (i.e., iLIR [19] and hfAIM [20]) as well as the recently described pLIRm software [22], a new computational tool for discriminating functional LIR motifs, which is provided as open-source Python code. In addition, there exist dedicated databases of predicted LIR motifs for particular species (i.e., iLIR database [23] and iLIR@viral [24]) which are easy for users to navigate and retrieve relevant predictions.

In the above-mentioned databases, the available LIR motifs are taxonomically restricted (e.g., the iLIR database provides predictions for 8 model species only), thus for species not covered by these resources, interested users need to directly use the available prediction tools. An additional source of data relevant to LIR motifs has become available with the most recent version of the ELM database of eukaryotic linear motifs [18] and its associated resource articles.ELM [25].

Summarizing, despite the increasing availability of in silico resources for LIR motifs, a number of inherent limitations exist. With respect to prediction tools, all currently available methods are unable to detect atypical cases of LIR motifs (e.g., non-canonical LIR motifs), while pLIRm is only available through the command line, thus preventing easy use by wet biologists. Moreover, the small size of available datasets has hindered the extensive validation of these prediction methods: in lieu of extended datasets of experimentally characterized LIR motifs, all prediction methods are not extensively/independently validated and it is possible that they suffer from overfitting [22]. Conversely, relevant databases come with their drawbacks too: iLIR database/iLIR@viral provide predicted LIR motifs only and in a taxonomically restricted manner, while ELM provides experimentally verified instances of LIR motifs but with small coverage only (67 LIR-motif instances under four different ELM classes: ELME000368–ELME000371; last accessed 15 May 2022).

An alternative – yet more cumbersome – route to obtain LIR-motif related information supported by experimental evidence is to directly search within the PubMed database [26] or other literature databases, e.g., Europe PMC [27], PubMedCentral [28]. This approach suffers from the fact that relevant information is scattered in a non-standardized manner within hundreds of publications, thus making it difficult to extract relevant data. For example, a PubMed search with the keyword-based query

```
"autophagy AND ('lir motif' OR 'aim motif'
OR 'lir-motif' OR 'aim-motif')"
```

(hereinafter “LIRquery”) identified 99 entries (query last performed on 15 May 2022) (Table S1). Of these, only 60 (60.7%) presented primary characterization data of at least one LIR motif while the remaining 39 (39.3%) are review papers or editorials on relevant topics. While reviews are unquestionably important in summarizing the literature in any particular topic, they more

often contain high-level information and readers wanting to obtain detailed information on particular LIR motifs (e.g., their originating sequences or their exact positions therein) need to eventually consult the primary research paper that describes the characterization of the motif.

According to our experience, several papers reporting the experimental validation of LIR motifs often deal with broader topics, thus motif characterization and relevant information may not be directly reflected in the paper title or abstract, which is where the PubMed search engine is searching for the query keywords. When performing the “LIRquery” against the full-text literature databases EuropePMC and PubMedCentral we retrieved 899 and 1150 entries respectively (queries performed on 24 May 2022). In such long lists, it is expected that several papers will only be marginally relevant to the topic of interest (e.g., reviews); many more are completely irrelevant for finding information confirming the functionality of particular LIR motifs, with a range of cases where LIR motifs are thoroughly reviewed or simply mentioned. For instance, when examining the 20 most recent entries retrieved from EuropePMC with the “LIRquery” only 3 entries (15%) corresponded to papers describing the characterization of LIR motifs (Table S2). While this dataset can by no means be considered representative of the underlying body of literature, it gives a first glimpse of the number of false positives returned by similar full-text queries. Carefully studying the text of these literature entries (and often extensive supplementary files) to extract information about LIR motifs requires tremendous effort, even from experienced researchers. A collection of literature-supported evidence for functional LIR motifs should provide interested researchers with easy and quick access to data, available in a standardized form, thus facilitating additional research efforts for deciphering the mechanisms of selective macroautophagy and related processes.

Herein, we describe our efforts on the creation of a comprehensive collection of experimentally characterized LIR motifs based on extensive biomedical literature search, aided by simple text analysis techniques, and followed by meticulous manual curation of full text papers. Furthermore, we present the development of a web accessible database (LIRcentral; <https://lircentral.eu>) in which interested researchers are able to retrieve information about literature-validated LIR motifs and display them along with features annotated in UniProt [29]. By cross-referencing protein entities to UniProt entries, LIRcentral enables seamless data integration with other resources. We also provide example cases where using LIRcentral can enable posing and answering interesting questions related to LIR motifs and LIRCPs in reasonable time, even for users with limited computational skills/background.

Results and Discussion

Content of the LIRcentral database

Curated publications in LIRcentral: As of writing this text (12 June 2022) a total of 159 peer-reviewed manuscripts from 56 different journals (Table 1) have served for the curation of LIRcentral entries (See Materials and Methods). Even though information regarding LIR motifs increasingly often

Table 1. Sources of journal papers ($n = 159$) reporting the experimental validation of LIR motifs currently curated for LIRcentral entries.

Journal title	# papers	Journal title	# papers
Autophagy	24	Cell	2
The Journal of Biological Chemistry	14	Cell Death & Disease	2
Nature Communications	12	Cell Metabolism	2
eLife	9	Cell Reports	2
The Journal of Cell Biology	9	Communications Biology	2
Molecular Cell	8	Current Biology	2
Developmental Cell	6	FASEB Journal	2
EMBO Reports	5	PLoS One	2
Nature Cell Biology	5	PNAS	2
Nature	4	Science Signaling	2
Cell Death and Differentiation	3	The Plant Cell	2
Oncotarget	3	Other (Journals with 1 paper each)	32
The EMBO Journal	3		

Note: Papers describing LIR motifs get published in both niche and more generic journals. More than half of the total number of papers is published in the top 7 journals (in terms of relevant papers published) in this list.

appears in pre-prints or other unreviewed sources of information, we deliberately do not include these motifs in LIRcentral. Instead, we mark LIR motifs reported in such sources as candidate entries and wait until the work appears in a journal after peer-review for inclusion in LIRcentral. Thus, entries currently in LIRcentral were supported by at least one primary research paper which was manually curated for each LIR motif. Review papers served both as an additional means to track the primary literature as well as to resolve any ambiguities therein.

Properties of the LIRcentral literature corpus: We attempted to qualitatively compare the content of the abstracts of these 159 documents in contrast to a background composed from the abstracts of 1,000 manuscripts retrieved from PubMed using the keyword “autophagy”. By constructing word-clouds, some differences in the prevailing terms in these two (not necessarily disjoint) sets of texts became evident: the manuscripts used to annotate LIR-motif entries were enriched in terms such as “binding”, “protein”, “LC3”, “Atg8”, “interaction”, “selective”, reflecting the role of the LC3-LIR-motif interaction in selective macroautophagy, while the background set was enriched in more generic terms, e.g., “induced”, “expression”, “treatment”, “apoptosis” (Figure 1).

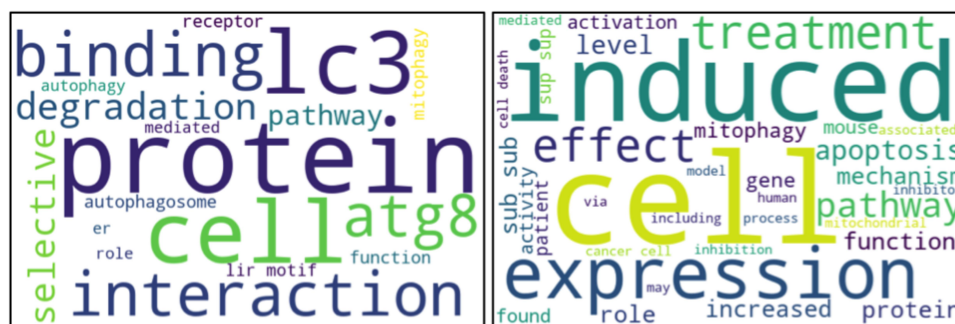
Comparison of the LIRcentral literature set to other (curated) collections: We compared the sets of curated papers used to populate LIRcentral to other relevant sets of publications: (i) the dataset of publications associated with the LIR-motif instances in the ELM database ($n = 57$; last accessed 9 May 2022) [18], (ii) the list of papers listed to be relevant to LIR motifs within the articles.ELM resource ($n = 103$; last accessed 9 May 2022) [25], and (iii) the manuscripts linked to instances of LIR motifs used to train/validate pLIRm ($n = 46$; taken from Supplementary Data 1 in [22]). We also used the results of the “LIRquery” against PubMed (accessed 9 May 2022) as a control.

From the Venn diagram representation (Figure 2) it was obvious that these collections were largely heterogeneous. Only 1 paper was common across all 5 collections: this is the seminal paper on the topic by Alemu and colleagues [15].

The LIRcentral literature collection was the most comprehensive and contained the most unique elements ($n = 80$) not available in any other literature dataset. These 80 papers, uniquely available through LIRcentral, span from 2009–2020 and mainly report primary results (only 6 marked as “Review” in PubMed). On the other hand, a total of 153 papers in this collective literature corpus were not curated in the current LIRcentral version of which (i) a large fraction (58, ~38%) have been published during the last 3 years, and (ii) 21 (~14%) have been marked in PubMed as review papers.

The articles.ELM resource contained a large number of unique elements ($n = 60$) several of which were either false positive predictions – i.e., they do not describe experimental validation of LIR motifs – (e.g., papers with PubMed identifiers [PMIDs]: 30078701, 28388439, 27422009, 27594685) or deal with other important functional aspects of LIR motifs, including post-translational modifications (e.g., PMID: 27757847) or the higher order structural organization and interactions of LIRCPs (e.g., PMID: 25921531, 26506893); however, it is worth noting that this set was assessed without taking into account the confidence score accompanying each publication.

Even though we had not manually checked all papers in detail, from the 5 unique ELM literature citations two were apparently unrelated to LIR motifs (PMID: 24668265, 24668266). These were used for annotating motif instance ELMI004471 in human WD repeat and FYVE domain-containing protein 3 (WDFY3/ALFY; UniProt: Q8IZQ1) and

**Figure 1.** Word cloud representation of the textual content of all abstracts from the papers curated in the current LIRcentral version (left) versus 1000 articles retrieved by the keyword “autophagy” using the Bio:Entrez module of Biopython (right).

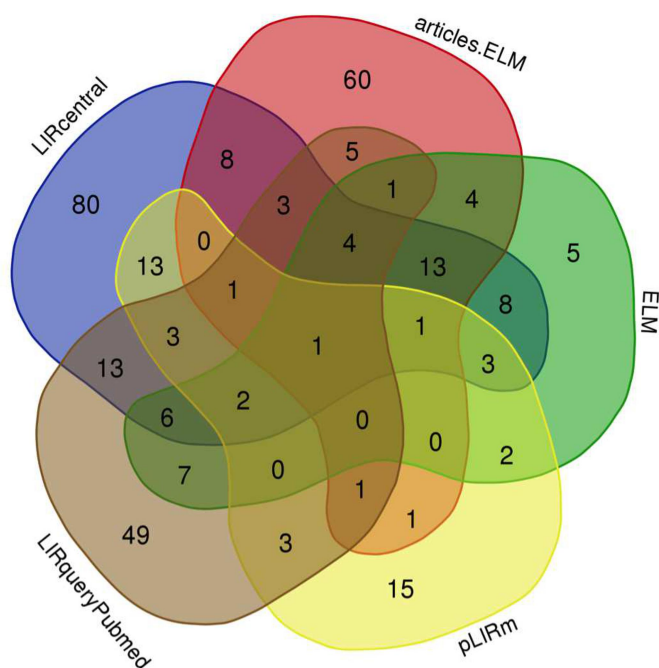


Figure 2. A Venn diagram representation of the overlaps between different literature collections related to LIR motifs. Drawn using <https://bioinformatics.psb.ugent.be/webtools/Venn/>.

were communicated to the ELM curators for correction (11 June 2022); according to the ELM development team these literature entries will be corrected in a next update of ELM (Manjeet Kumar and Toby Gibson, personal communication).

There were also 15 papers appearing in the pLIRm collection only. These were published between 2014 and 2020, with 3 entries marked by PubMed as “Review”. We will delve deeper into the pLIRm collection in a separate work (currently in preparation), since this literature dataset was used to annotate instances of LIR motifs which were subsequently used to train and validate computational methods presented in [22].

The distribution of publication dates for manuscripts describing experimentally verified LIR motifs illustrated an increasing overall interest for the study of LIR-motif mediated protein-protein interactions and the progress in the field (Figure 3). There was an overall increasing trend in the manuscripts used to annotate LIR motif entries in LIRcentral up to 2017 (included). The relative decline in content published after 2018 is due to the fact that (a) our massive semi-automatic analysis of full-text articles (see Materials and Methods) covered the period up to November 2020 and (b) recently, we concentrated our efforts on developing and making available the online LIRcentral database, with a backlog of dozens of papers/LIR motifs that are in the pipeline for inclusion in the next major release (scheduled to appear in Spring 2023). In any case, LIRcentral consistently covered comparable or (frequently) more papers to the other resources across the complete range from 2007 to date.

It is worth mentioning that while developing LIRcentral, we initially focused on covering as many LIR motifs as possible, without spending additional resources on covering all relevant literature. That is, when a newly published paper provided additional information on the functionality of a particular motif already annotated using another primary work, we intentionally left it aside for future curation, in order to curate papers related to motifs not currently in the database.

LIR-motif entries: The LIRcentral database follows a motif-centric approach, in the sense that LIRcentral entries correspond to instances of motifs (rather than protein/polypeptide entries or publications). The current LIRcentral version is composed of 292 instances of experimentally verified motifs, contained in 160 unique protein chains, originating from 22 species. Information in published manuscripts was used to identify the respective UniProt entries, and in some cases particular isoforms when this information could be deduced from the underlying curation process. LIRcentral

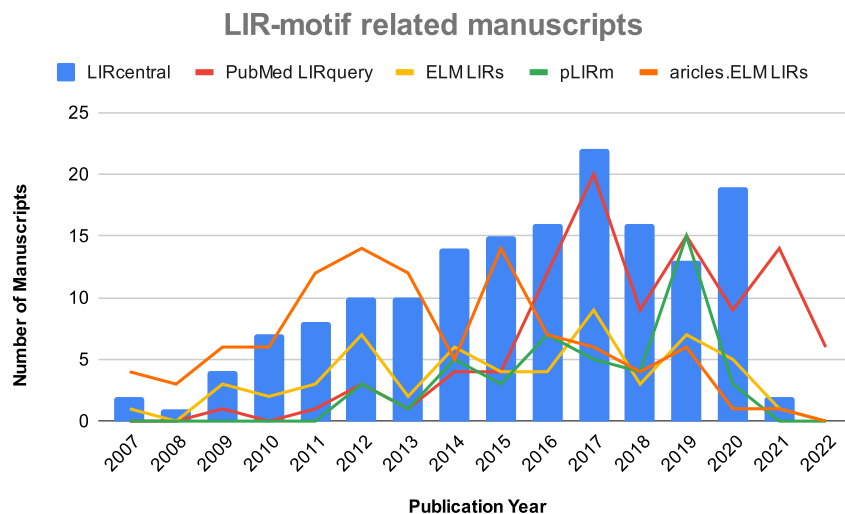
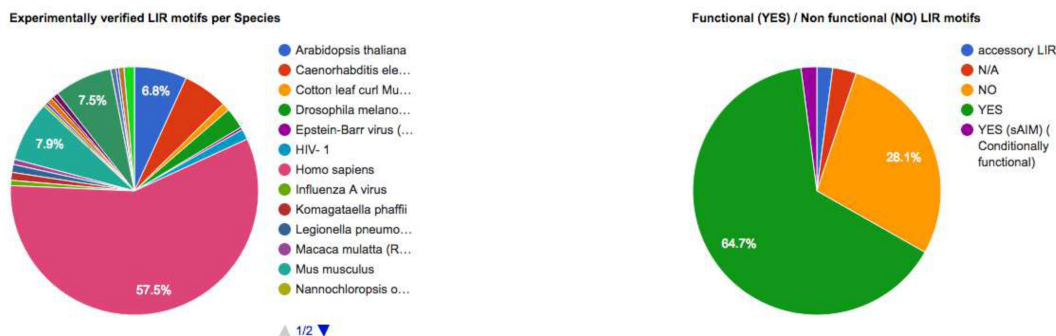


Figure 3. Distribution of publication dates for manuscripts describing experimentally verified LIR motifs in the current version of LIRcentral and comparison to other literature collections of LIR motifs.

Members of the Atg8 protein family (1) are small proteins with a ubiquitin-like fold and are key components of the macroautophagy machinery, participating essentially in all steps of macroautophagy from phagophore nucleation, to autolysosome formation (2). They are involved in an increasing number of biological processes, including LC3-associated phagocytosis (3, 4), endosomal microautophagy (4, 5), and unconventional secretion (6). In selective macroautophagy, selectivity is mainly achieved through receptor proteins that bind to the surface of Atg8 by anchoring a small linear peptide (AIM in plants and fungi, LIR-motif in mammals) (7). After the discovery of the first few LIR-containing proteins (LIRCPs), several definitions of the LIR-motif have appeared in the literature, highlighting a short core consensus sequence described by the regular expression pattern [WFY]xx[VLI] (8, 9, 10, 11, 12). Accumulating experimental evidence pinpoints the importance of regions flanking the core LIR-motif, while several instances of atypical LIR-motifs and LIR-independent interactions are being uncovered. We have developed LIRcentral as a publicly available, comprehensive collection of experimentally confirmed LIR-motifs, based on extensive biomedical literature search, assisted by text mining techniques and careful manual curation. Interested researchers are able to retrieve information about literature-validated LIR-motifs and display them along with sequence features annotated in UniProt. A pre-print describing LIRcentral is available in bioRxiv.



Bioinformatics Research Laboratory



LIRcentral is co-funded by the European Union (European Regional Development Fund, ERDF) and the Republic of Cyprus through the project EXCELLENCE/0421/0576 under the EXCELLENCE HUBS programme of the Cyprus Research and Innovation Foundation.

University of Cyprus

Figure 4. The LIRcentral landing page (“Home”). It provides a short introduction to the purpose of LIRcentral with a brief summary of the database content and links to all LIRcentral functionalities.

recorded both functional (i.e., binding; $n = 201$) and non-functional ($n = 82$) instances of LIR motifs, since negative examples can also be of great value, especially when developing prediction methods.

For the purposes of LIRcentral we identify three subclasses of binding LIR motifs:

- (i) Generic (labeled under the “Experimentally verified” field as “YES”).
- (ii) Conditionally functional, non-canonical shuffled LIR motifs – labeled as “YES (sAIM) (Conditionally functional)”. This subclass was introduced to describe atypical LIRs in human CDK5 regulatory subunit-associated protein 3 and its *Arabidopsis thaliana* homolog (UniProt: Q96JB5, Q9FG23 respectively [30]): such motifs are required for the binding of another functional LIR motif, even though they cannot independently maintain the interaction.
- (iii) Accessory LIR motifs, which function complementary to other functional (or accessory) LIR motifs, as for example the two non-canonical LIR motifs found in yeast Atg19 [31], which complement a canonical C-terminal motif and might play a role in defining the strength or the stoichiometry of the LC3-LIRCP interactions.

As verified but nonfunctional LIR motifs (labeled as “NO”) we include those cases that have been unambiguously screened (e.g., via mutating the critical X_0 , X_{+3} positions of the motif) to not bind Atg8s. In several cases we can safely

assume that all other peptides besides those directly shown to be functional LIR motifs can be considered as negative examples (i.e., nonfunctional). Nevertheless, at this point we decided to leave them unlabeled, since there are examples in the literature when revisiting a protein with characterized functional LIR motifs provides updated information that is not coherent with such an inference (see for example the history of the discovery of functional LIR motifs in yeast Atg19 [31,32]).

In addition, there are several cases of proteins where it might be known that they interact with members of the Atg8 family in a LIR-dependent manner, yet the underlying binding motif(s) have not been characterized so far (e.g., several proteins studied in [33]). LIR motifs predicted in such cases are designated as N/A (and can be interpreted as “candidate” functional motifs), since we have not found literature that experimentally demonstrates which of the predicted motifs are indeed functional.

Furthermore, for all proteins with an experimentally validated LIR motif, LIRcentral makes available all peptides matching the following regular-expression based definitions: canonical LIR motif ([WFY]xx[VLI]), the extended LIR motif/xLIR [19], and the five different motifs implemented in the hfAIM method [20]. Such motif instances are designated as “Predicted” and are not included in the above statistics.

The vast majority of LIRcentral entries comprised human LIR motifs (168, 57.5%) followed by entries from model species such as *Saccharomyces cerevisiae* (22, 7.5%), mouse (23, 7.9%) and *Arabidopsis thaliana* (20, 6.8%). It is worth mentioning that within LIRcentral entries there were also a few LIR motifs of non-eukaryotic origin.

The screenshot shows the LIRcentral search interface. At the top, there are navigation links: Home, Search, Browse, The LIRcentral Team, and Help. Below this is a search filter section with radio buttons for 'Verified (Reviewed / Under review)', 'All', and 'Predicted'. There are two input fields for species names: 'Toxoplasma' and 'Plasmodium'. A 'Submit' button is located below the input fields. Below the search section, there is a table of results with columns: Species, UNIPROT ACC, UNIPROT ID, Protein name, Motif type, and Up-stream. The table contains 5 entries. Below the table, there are buttons for 'COPY', 'EXCEL', 'CSV', and 'PDF'. At the bottom, there is a footer with logos for the European Union, the University of Cyprus, and the LIRcentral project.

Species	UNIPROT ACC	UNIPROT ID	Protein name	Motif type	Up-stream
<i>Plasmodium berghei</i>	A0A0Z0A9Q8	A0A0Z0A9Q8_PLABE	Early transcribed membrane protein	Non canonical LIR	SNEQHKYLIN
<i>Plasmodium falciparum (isolate 3D7)</i>	C0H519	C0H519_PLAF7	Autophagy-related protein 3, putative	Non canonical LIR	VHDLKIVDND
<i>Plasmodium falciparum (isolate 3D7)</i>	Q8IEU1	Q8IEU1_PLAF7	Early transcribed membrane protein 13	Non canonical LIR	SNDQKDSHVN
<i>Toxoplasma gondii (strain ATCC 50853 / GT1)</i>	S7URH8	S7URH8_TOXGG	Putative autophagy-related protein 3 atg3	F LIR	LRDDVFDLIN
<i>Toxoplasma gondii (strain ATCC 50853 / GT1)</i>	S7URH8	S7URH8_TOXGG	Putative autophagy-related protein 3 atg3	Non canonical LIR	TRVGHDEVEGG

Figure 5. Advanced search via the “Browse” functionality. Here all LIRcentral experimentally verified entries originating from species whose names match either “Toxoplasma” or “Plasmodium” are retrieved in a single query.

LIRcentral access and functionality

Users can get access to LIRcentral with any modern web browser pointed to the URL <https://lircentral.eu> (Figure 4). A simple “Search” functionality facilitates entry retrieval using keyword-based queries, while the “Browse” functionality offers a more advanced search interface where complex, targeted queries can be performed. Results retrieved after a successful “Search” or “Browse” operation are displayed in tabular form (Figure 5); these tables can be copied (as text in the clipboard) or exported to spreadsheet, comma-separated or PDF formatted files. Information on how to use LIRcentral is offered in a detailed online “Help” page (<https://lircentral.eu/help>). Users can easily download the complement of LIR-motif entries available in LIRcentral by performing a “Search” operation without any keywords. By selecting the appropriate radio button, users may select to retrieve motifs with experimental validation – “Verified (Reviewed/Under review)” option, motif instances predicted to match simple regular expression definitions of LIRs (“Predicted” option), or all motifs (“All” option).

Each protein with at least one experimentally characterized LIR motif entered in LIRcentral has its own dedicated page. There, all validated as well as predicted LIR motifs are displayed in tabular form. A graphical illustration of LIR motifs accompanied with features retrieved “on the fly” from the UniProt database is also provided.

Case studies

We anticipate that LIRcentral will be used by different researchers in tackling a wide range of research questions and applications. The most straightforward use of LIRcentral will be in cases where a researcher will be interested in transferring knowledge about a LIR motif described in the literature to a protein homolog in another species; retrieving reliable data from LIRcentral and combining them with readily available sequence analysis tools can guide the first steps for characterizing LIR motifs in homologs. In the following, we present examples of two “larger-scale” case studies, which can greatly benefit from the capabilities offered by LIRcentral and the knowledge collected therein.

Case study 1 - Revealing defining sequence properties of functional LIR motifs

Q: Which sequence properties distinguish a functional from a nonfunctional LIR motif?

This is an example of a valid (and, admittedly, interesting) biological question that researchers working in the field might want to address using LIRcentral (undoubtedly, we could imagine many more questions of a similar nature!). We present a step-by-step guide on how, using LIRcentral in

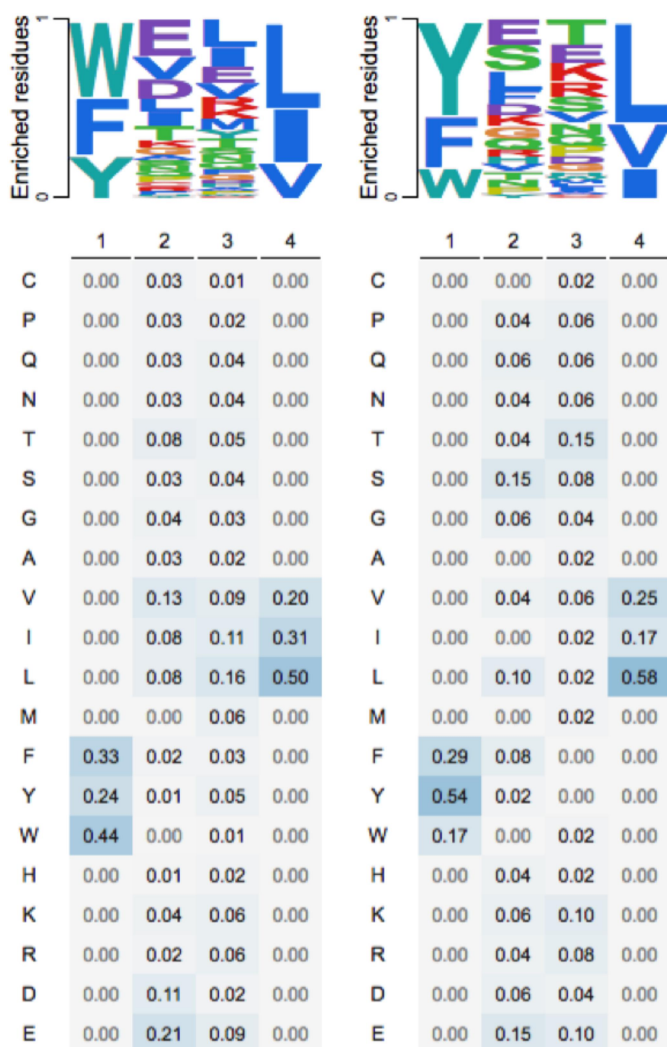


Figure 6. Positional frequencies of different amino acid residue types for human experimentally verified functional (left; $n = 101$) and nonfunctional (right; $n = 48$) sequences matching the canonical LIR motif. Top: Logo-like representations; Bottom: Actual frequencies. Computations performed using PSSMsearch [36] using experimentally verified human LIR motif data retrieved from LIRcentral.

combination with online tools, a researcher with little bioinformatics expertise could address the above question.

Ideally, it would be desirable to obtain sets of experimentally determined LIR motifs that belong either to the functional or the nonfunctional class. To avoid any unwanted inhomogeneity that might occur due to different species-specific compositional biases [34,35] we will focus on human LIR motifs only. In addition, all instances of non-canonical LIR motifs are excluded from the analysis.

To obtain all human LIR motifs with experimental support, a user can navigate to the “Browse” page and execute a targeted search, selecting the “Verified” radio button, to restrict entries retrieved among only those which are annotated based on literature curation. By selecting the “Species” option from the drop-down menu, users can enter “Homo sapiens” as the target species and execute the search. The results of this query (168 motif entries) can be exported in a spreadsheet file (using the “EXCEL” export button) and can be easily filtered to obtain subsets fulfilling the above criteria

within minutes. For simplicity, in this case-study we removed all non-canonical human LIR motifs and used the remaining 149 motif instances (101 functional, 48 nonfunctional) as input to the PSSMsearch server [36] for quickly computing positional frequencies for all residue types for the core LIR motif (Figure 6).

In fact, this approach (with some methodological modifications) has already been used to develop part of a new computational scheme to predict functional LIR motifs in *Homo sapiens* proteins (Chatzichristofi et al., in preparation). In particular, we propose the use of two PSSM models computed to represent functional and nonfunctional instances of human LIR motifs along with their flanking regions, combined with the average pLDDT confidence values from AlphaFold2 predictions [37] as a proxy for the “disorderliness” of these segments [38]. These results can then be combined using a machine learning classifier for accurate prediction of human LIR motifs.

Case study 2 - Constructing a protein-protein interaction-driven, LIRCP network in *Saccharomyces cerevisiae*

With the rapid advancement of -omics methods, the identification of new protein sequences is happening at an exponential rate. However, assigning biological function to a protein is still a challenging problem which often requires years of experimental work. Protein-protein interaction (PPI) networks are becoming increasingly popular, since they facilitate the enhanced description and understanding of cellular systems in physiological conditions and disease [39]. They can also aid in deciphering the biological function of proteins [40] and the prediction of protein complexes [41]. Additionally, PPI networks can provide useful insights on potential interacting partners, enabling experimental biologists to prioritize research targets. In this case study, we demonstrate how to efficiently generate a PPI network centered around LIRCPs, using a list of manually curated LIRcentral entries belonging to *S. cerevisiae* (baker’s yeast).

Users can easily retrieve a list of yeast protein accessions by using the LIRcentral “Browse” functionality, searching for the keyword “cerevisiae” under the “Species” field after selecting the “Verified” radio button. This action will result in displaying a table containing all yeast entries with at least one experimentally characterized LIR motif (22 characterized motifs, of which 19 functional in 12 different proteins). The entries can then be downloaded in any of the available formats (e.g., csv, excel). For simplicity, we downloaded results in “excel” format. Using spreadsheet-like tools to further process the data, users can easily extract the list of UniProt accessions and provide them as input to tools like GeneMANIA [42] to generate a PPI network focusing on yeast LIRCPs within just a few minutes (Figure 7).

In particular, we provided the 12 yeast LIRCPs as input and asked GeneMANIA to enrich this collection with 20 proteins based on experimentally determined PPIs – an obvious GeneMANIA “finding” being Atg8. Among the proteins retrieved from this analysis were Sfb3, a component of COPII vesicles, which is known to act with Atg40 in reticulophagy [43]. Another protein identified by GeneMANIA is Trs33,

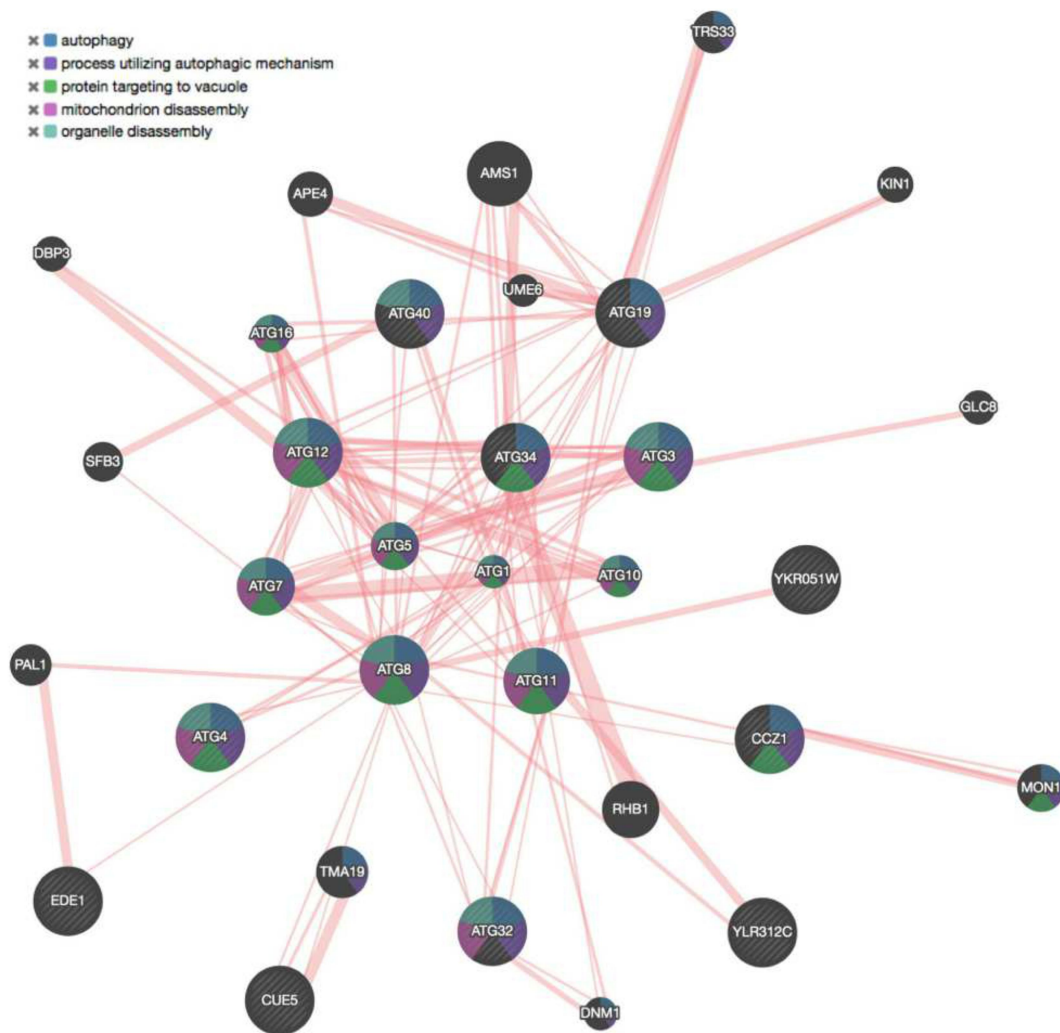


Figure 7. PPI network using as input a list of 12 verified yeast LIRCP entries from LIRcentral. Proteins are represented as nodes while edges represent physical interactions between them. The input set was enhanced with the prediction of 20 additional entities based on known PPIs. Striped nodes signify input proteins and the color scheme corresponds to the top 5 gene ontology terms ($FDR < 10^{-12}$). This network was generated using the GeneMANIA web server <https://genemania.org> [42]; only yeast PPI data were used to infer 20 proteins related to the input dataset (all other parameters set to their default values).

a subunit of the TRAPPIII and TRAPPIV complexes which are required for Ypt1-mediated autophagy [44]. An interesting case in this list is Pal1, which was initially characterized as an early-arriving endocytic coat protein in yeast via its interaction with Ede1 [45], an interaction independently confirmed more recently [46]; in subsequent work a high-throughput screen showed Pal1 to physically interact with Atg8 in the context of a selective macroautophagy pathway for protein condensates formed by proteins involved in clathrin-mediated endocytosis [47]. Pal1 contains several peptides matching canonical LIR motifs, however whether the latter interaction is LIR-motif mediated needs to be further validated. The above examples illustrate that, using data from LIRcentral, researchers can become aware of possible cross-talk of autophagy and its components to other cellular processes.

Conclusions

We anticipate that LIRcentral will become a key resource for researchers requiring access to reliable information regarding

LIRCPs and their interactions with Atg8 homologs. A small, but dedicated, group of curators have made this first version of LIRcentral possible by collecting relevant papers and meticulously extracting information relevant to the experimental characterization of LIR motifs. This information is currently impossible to be automatically extracted from papers scattered across different publication media in a non-structured manner, but it can be invaluable at the hands of both wet and computational biologists for addressing a range of research questions related to selective macroautophagy or other cellular processes where Atg8 family members participate. We have deliberately chosen to refrain from handling the cases of LIR-independent modes of binding to LC3 (e.g., with ubiquitin-interacting motifs, UIMs [48]) to maintain our focus on those interactions happening at the particular LC3-LIR-motif interface: in that sense, LIRcentral already contains several atypical LIR-motif cases.

With the advancement of our collective knowledge in the details of the underlying interactions, we expect to be able to meaningfully expand LIRcentral in future releases. For example,

in mammalian species a specific subclass of LIR motifs seems to show particular specificity toward the GABARAP clade of Atg8 homologs (GABARAP interaction motifs, GIMs [49]). This specificity seems to be reflected in the underlying sequences at the core motif but also in the flanking regions. Hopefully, we will be able to include this important piece of information within the next major release of LIRcentral. Other possible expansions, with respect to the type of information recorded and made available with each LIR motif, include context-specific functionality (e.g., LIR-motif switches based on post-translational modifications) and inclusion of information related to the experimental methods used for characterizing LIR-motif functionality.

In addition, we believe LIRcentral already offers a simple and intuitive interface for data retrieval. However, we are aware that certain improvements can further enhance the user experience with this resource. A functionality we intend to implement for providing alternative or more efficient ways to use LIRcentral is, for example, the possibility to support search based on sequence similarity.

One of the main obstacles the LIRcentral group needs to overcome is the difficulty to sustain funding for database curation. For this purpose, we have planned to develop a mechanism (including software and database components) that will enable participation of any interested parties from the research community in the literature curation process (i.e., “crowdsourcing”).

This community effort will not only help in keeping the content up to date with the most current literature but will also help in developing community-supported standards (e.g., how should LC3-LIR-motif interaction data be reported?) and propose solutions to difficult, “strategic” problems. One example of the latter stems from the current dependence of LIRcentral from UniProt. With the foreseen possibility that an increasing number of synthetic/engineered LIR-motif containing peptides will be available in the literature (as for example in [50]), there needs to be decided whether such entries should be cataloged in LIRcentral (and how).

As a first step to facilitate community-driven efforts in the context of LIRcentral, we have already partnered with the APICURON resource (<https://apicuron.org/> [51]), which provides the means to monitor and expose the curation efforts of individuals, thus giving appropriate credit to all researchers who put effort in providing new or revising existing LIRcentral content.

Similar to other cases of SLiMs, LIR motifs are often located within intrinsically disordered regions [16–18], a notion that has already been exploited in reducing false positive hits in LIR-motif detection tools [19]. Recognizing the importance of this association, members of the LIRcentral team participated in the preparation of the “Autophagy collection” for DisProt release 2022_06 (<https://disprot.org/>, released online on 28 June 2022), where we collaborated in the annotation of known functional LIR motifs residing in experimentally verified intrinsically disordered regions [52]. Interestingly, we confirmed that only for a few cases of LIR motifs there exists hard experimental evidence of intrinsic disorder: there were only 18 such entries in DisProt mostly from human proteins (10) highlighting the necessity to collect more data across different species.

Currently, an active group of curators provides content, while regular updates to LIRcentral are planned (approximately every six months). We open a call to the autophagy community for

contributing in the curation effort, for example by submitting information about newly discovered (or updating information on known) LIR motifs. We anticipate that such a “crowdsourcing” of the curation process among the active autophagy research community members will (i) help toward increasing both quality and coverage of the curated data, (ii) promote the development of community-agreed ways of reporting relevant data and (iii) stimulate further research in the area of selective macroautophagy and engage younger researchers in the field, while at the same time it will facilitate the long-term sustainability of LIRcentral.

Materials and Methods

Defining a strategy for collecting relevant literature

In order to be able to create a high-quality dataset of experimentally verified LIR motifs for inclusion in LIRcentral we had to devise a strategy for retrieving as many relevant published papers as possible to ensure high coverage. In addition, due to the significant amount of time necessary to carefully process papers for curation, a minimum number of false positives is desirable, since the limited curation resources available should not be wasted on working with irrelevant papers.

We started with an initial collection of papers, mostly retrieved by direct PubMed search (keyword-based query, i.e., using the “LIRquery”) or reverse search, i.e., papers citing key papers on LIR motifs (e.g [15]) or the software tools/database for LIR motif prediction (e.g [19,20,23]). Papers relevant to the description of LIR motifs that could not be retrieved by the “LIRquery” were manually inspected to identify why they are unidentifiable using direct search.

We identified several reasons why papers describing the experimental characterization of LIR motifs do not contain a relevant mention in the title or abstract that could facilitate their automatic retrieval, the most prominent being:

- (1) Older papers – when the terms LIR/AIM motif were not yet coined or widely established (PMIDs: 17580304, 17916189, 18027972, 19021777).
- (2) Papers where LIR-motif characterization might not be central in the paper’s story (e.g., PMIDs: 29937374, 29277967, 28244873), where this information is “buried” in the main text, in figure legends or even in supplementary data.
- (3) Papers mentioning LIR motifs in a different number of (non-standardized) ways.

As an example demonstrating the latter case, we can see several different ways of expressing the concept of LIR/AIM motifs in the literature:

- “LC3-interacting regions (LIRs)” (PMID: 29717061)
- “LC-3 interacting regions (LIRs)” (PMID: 30568238)
- “light chain 3 (LC3)-interacting region (LIR) motifs” (PMID: 29937374)
- “LC3-interaction region domain” (PMID: 29416008)
- “Atg8-family-interacting motif” (PMID: 30451685)
- “Atg8 family-interacting motif” (PMID: 30026233)
- “Atg8-binding sequence motif” (PMID: 29867141)

Therefore, it became obvious that to retrieve the maximum possible number of papers for curating relevant information, we should find some simple rules based on the complete body of text included in publications, when available.

After curating a number of relevant papers, we observed that experimental evidence for the characterization of LIR motifs is most often described in sentences containing terms with the following characteristics:

- Describe entities that refer to an Atg8 homolog (e.g., “LC3”, “Atg8”) or a LIR motif (e.g., “LIR”, “AIM”), and
- A term (often a verb) indicating interaction (e.g., “interact”, “bind”, “recruit”), and
- A term suggesting a conclusion, signified by terms such as “confirm”, “identify”, “show”, “demonstrate” etc.

In addition, there are cases where the exact sequence of the motif discovered is mentioned.

Literature corpus generation and curation

Papers describing the experimental validation of LIR motifs are acquired using a combination of “bulk” (see below) and “selective” (direct and reverse PubMed search, see above) approaches.

For bulk retrieval, we first downloaded locally XML files for all papers with open access full text available in PubMedCentral matching the keyword “autophagy”. Such searches have been incrementally performed to cover the period up to November 2020, resulting in 78,798 publications. Then, we developed a simple text analysis pipeline in the Perl programming language, implementing the rules identified to be associated with sentences describing the experimental characterization of LIR motifs.

This pipeline was used to extract snippets, which are (parts of) sentences with the characteristics mentioned in the previous section: namely, (a) a term denoting an Atg8 homolog or a LIR motif, (b) a term signifying interaction, and (c) a term suggesting conclusion. Two example snippets with the relevant terms underlined are shown below:

[PMID:24528869] “[...] We conclude that the M2 cytosolic tail contains a conserved LIR motif capable of directly interacting with LC3 in infected cells. [...]”

[PMID:21893048] “[...] Here, we confirm this interaction and identify an Atg8 interacting motif (AIM) in Stbd1 necessary for GABARAP1 binding [...]”

Even though the above examples are clearly suggestive that the respective papers most likely describe the experimental characterization of LIR motifs, there are many other instances of irrelevant ones. All such snippets of text were exported in a spreadsheet-like file and highlighted articles worth investigating further. A total of 5151 papers with at least one informative portion of text were further manually inspected to identify papers for adding to the curation pipeline. Even though we did not formally validate the specificity and sensitivity of this approach for identifying manuscripts describing LIR-motif-LC3 interactions, we observed that it works well in practice (see also [Figures 2 and 3](#)). Also, despite a non-negligible number of false

positives still retained in the literature corpus to be manually examined by LIRcentral curators, the size of the corpus after this filtering was manageable.

All selected papers underwent a curation process following the buddy system: at least two curators studied each paper and used the information provided therein (including data supplements) to correctly identify:

- (i) the species corresponding to the protein analyzed,
- (ii) a UniProt identifier that unambiguously corresponds to the protein entity analyzed in each particular paper,
- (iii) experimental evidence for the characterization of candidate LIR motifs, for example mutation of key LIR motif residues followed by binding assays against Atg8s etc.
- (iv) the position of the core LIR motif. Importantly, often the authors in the literature may report different sequence spans as the motif region. In order to have a normalized view across different proteins, what we report in LIRcentral is not the author-based definition but the position of the core motif, unless we are dealing with atypical motifs. One such extremely atypical case is the split LIR motif of yeast Atg12 (UniProt: P38316), where two distal residues (Ile111, Phe185) mimic the structure of the hydrophobic and aromatic residues of a canonical LIR motif [53].

The curations of individual “buddies” were juxtaposed and underwent final corrections before they ended up in the database.

A graphical overview of the literature collection and curation steps followed is depicted in [Fig. S1](#).

Acknowledgements

The authors wish to thank former members of the Bioinformatics Research Laboratory at the University of Cyprus (especially Stelios Tsompanis) for earlier contributions to LIRcentral.

We also thank the APICURON team, in particular Dr Federica Quaglia (University of Padua), for fruitful discussions and efforts for integrating LIRcentral with APICURON. We also thank Elena Papaleo for constructive comments and discussions during the LIRcentral kick-off meeting in the University of Crete (April 28-29, 2022). We are also indebted to authors of papers describing LIR motifs who provided us access to full text when that was not available through our institutional library subscriptions.











Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work has been co-funded by the European Regional Development Fund and the Republic of Cyprus through the Cyprus Research and Innovation Foundation (Project: EXCELLENCE/0421/0576) and by the University of Cyprus [Computational approaches towards mechanistic insights and improved detection of functional LIR-motifs in selective autophagy receptors and adaptors. (idLIR)] internal grant (to V.J.P.).

ORCID

Agathangelos Chatzichristofi  <http://orcid.org/0000-0002-4365-654X>
 Vasileios Sagris  <http://orcid.org/0000-0001-6587-8357>
 Aristos Pallaris  <http://orcid.org/0000-0002-8125-5480>
 Marios Eftychiou  <http://orcid.org/0000-0002-5929-6956>
 Ioanna Kalvari  <http://orcid.org/0000-0001-9424-9197>
 Nicholas Price  <http://orcid.org/0000-0001-6672-2952>
 Theodosios Theodosiou  <http://orcid.org/0000-0002-6151-1803>
 Ioannis Iliopoulos  <http://orcid.org/0000-0002-9079-0565>
 Ioannis P. Nezis  <http://orcid.org/0000-0003-0233-7574>
 Vasilis J Promponas  <http://orcid.org/0000-0003-3352-4831>

References

- [1] Das G, Shrivage BV, Baehrecke EH. Regulation and function of autophagy during cell survival and cell death. *Cold Spring Harb Perspect Biol.* 2012;4:a008813. doi: 10.1101/cshperspect.a008813
- [2] Reggiori F, Mauthe M. At the Center of Autophagy: Autophagosomes. *Encyclopedia Of Cell Biology.* 2016;243–247. doi: 10.1016/B978-0-12-394447-4.20021-7
- [3] Johansen T, Lamark T. Selective autophagy: ATG8 family proteins, LIR motifs and cargo receptors. *J Mol Biol.* 2020;432(1):80–103. doi: 10.1016/j.jmb.2019.07.016
- [4] Heckmann BL, Green DR. LC3-associated phagocytosis at a glance. *J Cell Sci.* 2019;132:jcs.222984. doi: 10.1242/jcs.222984
- [5] Nieto-Torres JL, Leidal AM, Debnath J, et al. Beyond autophagy: The expanding roles of ATG8 proteins. *Trends Biochem Sci.* 2021;46(8):673–686. doi: 10.1016/j.tibs.2021.01.004
- [6] Mejlvang J, Olsvik H, Svenning S, et al. Starvation induces rapid degradation of selective autophagy receptors by endosomal microautophagy. *J Cell Bio.* 2018;217(10):3640–3655. doi: 10.1083/jcb.201711002
- [7] Claude-Taupin A, Jia J, Mudd M, et al. Autophagy's secret life: Secretion instead of degradation. *Essays Biochem.* 2017;61:637–647. doi: 10.1042/EBC20170024
- [8] Wesch N, Kirkin V, Rogov VV., et al. Atg8-Family Proteins—Structural Features and Molecular Interactions in Autophagy and Beyond. *Cells.* 2020;9:2008. doi: 10.3390/cells9092008
- [9] Pankiv S, Clausen TH, Lamark T, et al. P62/SQSTM1 binds directly to Atg8/LC3 to facilitate degradation of ubiquitinated protein aggregates by autophagy. *J Biol Chem.* 2007;282(33):24131–24145. doi: 10.1074/jbc.M702824200
- [10] Mohrlüder J, Stangler T, Hoffmann Y, et al. Identification of calreticulin as a ligand of GABARAP by phage display screening of a peptide library. *FEBS J.* 2007 Nov;274(21):5543–5555. doi: 10.1111/j.1742-4658.2007.06073.x
- [11] Mohrlüder J, Hoffmann Y, Stangler T, et al. Identification of clathrin heavy chain as a direct interaction partner for the gamma-aminobutyric acid type A receptor associated protein. *Biochemistry.* 2007 Dec 18;46(50):14537–14543. doi: 10.1021/bi7018145
- [12] Ichimura Y, Kominami E, Tanaka K, et al. Selective turnover of p62/A170/SQSTM1 by autophagy. *Autophagy.* 2008;4(8):1063–1066. doi: 10.4161/auto.6826
- [13] Noda NN, Ohsumi Y, Inagaki F. Atg8-family interacting motif crucial for selective autophagy. *FEBS Lett.* 2010;584(7):1379–1385. doi: 10.1016/j.febslet.2010.01.018
- [14] Noda NN, Kumeta H, Nakatogawa H, et al. Structural basis of target recognition by Atg8/LC3 during selective autophagy. *Genes Cells.* 2008;13(12):1211–1218. doi: 10.1111/j.1365-2443.2008.01238.x
- [15] Alemu EA, Lamark T, Torgersen KM, et al. ATG8 family proteins act as scaffolds for assembly of the ULK complex: Sequence requirements for LC3-interacting region (LIR) motifs. *J Biol Chem.* 2012;287(47):39275–39290. doi: 10.1074/jbc.M112.378109
- [16] Popelka H, Klionsky DJ. Analysis of the native conformation of the LIR/AIM motif in the Atg8/LC3/GABARAP-binding proteins. *Autophagy.* 2015;11(12):2153–2159. doi: 10.1080/15548627.2015.1111503
- [17] Popelka H. Dancing while self-eating: Protein intrinsic disorder in autophagy. *Prog Mol Biol Transl Sci.* 2020;174:263–305. doi: 10.1016/bs.pmbts.2020.03.002
- [18] Kumar M, Michael S, Alvarado-Valverde J, et al. The eukaryotic linear motif resource: 2022 release. *Nucleic Acids Res.* 2022;50(D1):D497–D508. doi: 10.1093/nar/gkab975
- [19] Kalvari I, Tsompanis S, Mulakkal NC, et al. iLIR: A web resource for prediction of Atg8-family interacting proteins. *Autophagy.* 2014;10(5):913–925. doi: 10.4161/auto.28260
- [20] Xie Q, Tzfadia O, Levy M, et al. hfAIM: A reliable bioinformatics approach for in silico genome-wide identification of autophagy-associated Atg8-interacting motifs in various organisms. *Autophagy.* 2016;12:876–887. doi: 10.1080/15548627.2016.1147668
- [21] Tóth D, Horváth GV, Juhász G. The interplay between pathogens and Atg8 family proteins: Thousand-faced interactions. *FEBS Open Bio.* 2021;11:3237–3252. doi: 10.1002/2211-5463.13318
- [22] Han Z, Zhang W, Ning W, et al. Model-based analysis uncovers mutations altering autophagy selectivity in human cancer. *Nat Commun.* 2021;12(1):3258. doi: 10.1038/s41467-021-23539-5
- [23] Jacomin AC, Samavedam S, Promponas V, et al. iLIR database: A web resource for LIR motif-containing proteins in eukaryotes. *Autophagy.* 2016;12(10):1945–1953. doi: 10.1080/15548627.2016.1207016
- [24] Jacomin AC, Samavedam S, Charles H, et al. iLIR@viral: A web resource for LIR motif-containing proteins in viruses. *Autophagy.* 2017;13(10):1782–1789. doi: 10.1080/15548627.2017.1356978
- [25] Palopoli N, Iserte JA, Chemes LB, et al. The articles.ELM resource: Simplifying access to protein linear motif literature by annotation, text-mining and classification. *Database (Oxford).* 2020;2020. doi: 10.1093/database/baaa040
- [26] Lu Z. PubMed and beyond: A survey of web tools for searching biomedical literature. *Database (Oxford).* 2011;2011:baq036. doi: 10.1093/database/baq036
- [27] Ferguson C, Araújo D, Faulk L, et al. Europe PMC in 2020. *Nucleic Acids Res.* 2021;49(D1):D1507–D1514. doi: 10.1093/nar/gkaa994
- [28] Beck J. “Report from the Field: PubMed Central, an XML-based Archive of Life Sciences Journal Articles.” Presented at International Symposium on XML for the Long Haul: issues in the Long-term Preservation of XML, Montréal, Canada, August 2, 2010. Proceedings of the International Symposium on XML for the Long Haul: issues in the Long-term Preservation of XML Vol. 6. Balisage Series on Markup Technologies. 2010. doi: 10.4242/BalisageVol6.Beck01
- [29] The UniProt Consortium. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res.* 2021;49:D480–D489. doi: 10.1093/nar/gkaa1100
- [30] Stephani M, Picchianti L, Gajic A, et al. A cross-kingdom conserved ER-phagy receptor maintains endoplasmic reticulum homeostasis during stress. *Elife.* 2020 Aug 27;9:e58396. doi: 10.7554/eLife.58396
- [31] Abert C, Kontaxis G, Martens S. Accessory Interaction Motifs in the Atg19 Cargo Receptor Enable Strong Binding to the Clustered Ubiquitin-related Atg8 Protein. *J Biol Chem.* 2016 Sep 2;291(36):18799–18808. doi: 10.1074/jbc.M116.736892
- [32] Shintani T, Huang WP, Stromhaug PE, et al. Mechanism of cargo selection in the cytoplasm to vacuole targeting pathway. *Dev Cell.* 2002 Dec;3(6):825–837. doi: 10.1016/s1534-5807(02)00373-8
- [33] Behrends C, Sowa ME, Gygi SP, et al. Network organization of the human autophagy system. *Nature.* 2010 Jul 1;466(7302):68–76. doi: 10.1038/nature09204
- [34] Mier P, Paladin L, Tamana S, et al. Disentangling the complexity of low complexity proteins. *Brief Bioinform.* 2020 Mar 23;21(2):458–472. PMID: 30698641; PMCID: PMC7299295. doi: 10.1093/bib/bbz007
- [35] Tsoka S, Promponas V, Ouzounis CA. Reproducibility in genome sequence annotation: the Plasmodium falciparum chromosome 2 case. *FEBS Lett.* 1999 May 28;451(3):354–355 PMID: 10371220. doi: 10.1016/s0014-5793(99)00599-2

- [36] Krystkowiak I, Manguy J, Davey NE. PSSMsearch: a server for modeling, visualization, proteome-wide discovery and annotation of protein motif specificity determinants. *Nucleic Acids Res.* 2018 Jul 2;46(W1):W235–W241. PMID: 29873773; PMCID: PMC6030969. doi: [10.1093/nar/gky426](https://doi.org/10.1093/nar/gky426)
- [37] Tunyasuvunakool K, Adler J, Wu Z, et al. Highly accurate protein structure prediction for the human proteome. *Nature.* 2021 Aug;596(7873):590–596. doi: [10.1038/s41586-021-03828-1](https://doi.org/10.1038/s41586-021-03828-1)
- [38] Piovesan D, Monzon AM. Tosatto SCE Intrinsic Protein Disorder, Conditional Folding and AlphaFold2 *BiorXiv.* 2022. doi: [10.1101/2022.03.03.482768](https://doi.org/10.1101/2022.03.03.482768).
- [39] Ideker T, Sharan R. Protein networks in disease. *Genome Res.* 2008 Apr;18(4):644–652. PMID: 18381899; PMCID: PMC3863981. doi: [10.1101/gr.071852.107](https://doi.org/10.1101/gr.071852.107)
- [40] Zhou J, Xiong W, Wang Y, et al. Protein Function Prediction Based on PPI Networks: Network Reconstruction vs Edge Enrichment. *Front Genet.* 2021Dec 14;12:758131. PMID: 34970299; PMCID: PMC8712557. doi: [10.3389/fgene.2021.758131](https://doi.org/10.3389/fgene.2021.758131)
- [41] Omranian S, Nikoloski Z, Grimm DG. Computational identification of protein complexes from network interactions: Present state, challenges, and the way forward. *Comput Struct Biotechnol J.* 2022May 27;20:2699–2712. PMID: 35685359; PMCID: PMC9166428. doi: [10.1016/j.csbj.2022.05.049](https://doi.org/10.1016/j.csbj.2022.05.049)
- [42] Warde-Farley D, Donaldson SL, Comes O, et al. The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Res.* 2010 Jul;38(Web Server issue):W214–20. PMID: 20576703; PMCID: PMC2896186. doi: [10.1093/nar/gkq537](https://doi.org/10.1093/nar/gkq537)
- [43] Cui Y, Parashar S, Zahoor M, et al. A COPII subunit acts with an autophagy receptor to target endoplasmic reticulum for degradation. *Science.* 2019Jul 5;365(6448):53–60. doi: [10.1126/science.aau9263](https://doi.org/10.1126/science.aau9263)
- [44] Lipatova Z, Majumdar U, Segev N. Trs33-Containing TRAPP IV: A Novel Autophagy-Specific Ypt1 GEF. *Genetics.* 2016 Nov;204(3):1117–1128. doi: [10.1534/genetics.116.194910](https://doi.org/10.1534/genetics.116.194910)
- [45] Carroll SY, Stimpson HE, Weinberg J, et al. Analysis of yeast endocytic site formation and maturation through a regulatory transition point. *Mol Biol Cell.* 2012 Feb;23(4):657–668. doi: [10.1091/mbc.e11-02-0108](https://doi.org/10.1091/mbc.e11-02-0108)
- [46] Moorthy BT, Sharma A, Boettner DR, et al. Identification of Suppressor of Clathrin Deficiency-1 (SCD1) and Its Connection to Clathrin-Mediated Endocytosis in *Saccharomyces cerevisiae*. *G3 (Bethesda).* 2019Mar 7;9(3):867–877. doi: [10.1534/g3.118.200782](https://doi.org/10.1534/g3.118.200782)
- [47] Wilfling F, Lee CW, Erdmann PS, et al. A Selective Autophagy Pathway for Phase-Separated Endocytic Protein Deposits. *Mol Cell.* 2020Dec 3;80(5):764–778.e7. doi: [10.1016/j.molcel.2020.10.030](https://doi.org/10.1016/j.molcel.2020.10.030)
- [48] Marshall RS, Hua Z, Mali S, et al. ATG8-Binding UIM Proteins Define a New Class of Autophagy Adaptors and Receptors. *Cell.* 2019Apr 18;177(3):766–781.e24. doi: [10.1016/j.cell.2019.02.009](https://doi.org/10.1016/j.cell.2019.02.009)
- [49] Rogov VV, Stolz A, Ravichandran AC, et al. Structural and functional analysis of the GABARAP interaction motif (GIM). *EMBO Rep.* 2018 Dec;19(12):e47268. doi: [10.15252/embr.201847268](https://doi.org/10.15252/embr.201847268)
- [50] Quinet G, Génin P, Ozturk O, et al. Exploring selective autophagy events in multiple biologic models using LC3-interacting regions (LIR)-based molecular traps. *Sci Rep.* 2022 May 10;12(1):7652. PMID: 35538106; PMCID: PMC9090809. doi: [10.1038/s41598-022-11417-z](https://doi.org/10.1038/s41598-022-11417-z)
- [51] Hatos A, Quaglia F, Piovesan D, et al. APICURON: a database to credit and acknowledge the work of biocurators. *Database (Oxford).* 2021 Apr 21;2021:baab019. PMID: 33882120; PMCID: PMC8060004. doi: [10.1093/database/baab019](https://doi.org/10.1093/database/baab019)
- [52] Quaglia F, Mészáros B, Salladini E, et al. DisProt in 2022: improved quality and accessibility of protein intrinsic disorder annotation. *Nucleic Acids Res.* 2022Jan 7;50(D1): D480–D487. PMID: 34850135; PMCID: PMC8728214. doi: [10.1093/nar/gkab1082](https://doi.org/10.1093/nar/gkab1082)
- [53] Kaufmann A, Beier V, Franquelim HG, et al. Molecular mechanism of autophagic membrane-scaffold assembly and disassembly. *Cell.* 2014 Jan 30;156(3):469–481. doi: [10.1016/j.cell.2013.12.022](https://doi.org/10.1016/j.cell.2013.12.022)