ATGHub Documentation

Index

Introduction **2**

1. Autophagy Databases 2

2. The Gene Ontology 2

Methods **4**

1. Database Implementation (ATGHub) 4

1.1. Data Collection *6*

1.2. Aggregation of Recurrence Data *6*

1.3. Functional Annotation *7*

1.4. Website Front-end Design *7*

Results **8**

1. In Silico Analysis: ATGHub 8

1.1. Identification of a Central Hub of Genes for Autophagy *8*

1.2. Web Interface for Dataset Querying *10*

Bibliography **11**

Websites **12**

Data Avaliability **12**

Introduction

1. Autophagy Databases

Given the complexity of the mechanisms involved in the autophagic process, publicly accessible web resources have been developed containing various types of information: peer-reviewed scientific articles (PubMed), functional annotations at the gene and protein level, molecular interactions, molecular structure, and relationships with metabolic pathways of genes involved in autophagy. The most relevant resource in this field is AutophagyNet v2.1 (https://www.autophagynet.org/), developed in 2022 by the Korcsmaros Group at the Earlham Institute (UK). This tool, based on the previous database known as ARN (Autophagy Regulatory Network), annotates genes based on the definition of a core of genes essential for the autophagic process. From the core, a multi-layer system was then constructed where each autophagic interactor was assigned to a layer and has various interactions with other layers and, above all, with the central core. The annotations in this resource are mainly based on experimental evidence (Csabai et al., 2024). Another relevant resource is HAMdb (http://hamdb.scbdd.com/), whose full name is Human Autophagy Modulator database, a database generated by integrating six other curated datasets on autophagy, also containing a list of chemical and protein modulators that regulate autophagy in humans (Wang et al., 2018). The latest important resource for studying autophagy processes in silico is HADB (https://www.autophagy.lu/), short for Human Autophagy Database. This database consists of an annotated set of autophagy genes, curated by the Luxembourg Institute of Health.

2. The Gene Ontology

Gene Ontology (Gene Ontology Resource) is an annotation system of gene function based on three functional characteristics: Molecular Function (MF), Cellular Component (CC), and Biological Process (BP). MF indicates the main activities performed by the gene at the molecular level. CC indicates in which cellular structure the MF are performed. BP indicates the set of molecular activities performed by the gene in order to achieve a specific result (Gene Ontology Consortium, 2023). Gene Ontology terms are identified by a unique code starting with “GO:” and are ordered hierarchically from a more general abstraction to a more specific one. An example to understand the order is the term *autophagy* (GO:0006914), whose ‘child’ terms are *macroautophagy* (GO:0016236), *microautophagy* (GO:0016237) and *chaperone-mediated autophagy* (GO:0061684). In turn, the term *autophagy* is part of *process utilizing autophagic mechanism* (GO:0061919). For

Biological Processes, the order of terms can also be determined by the type of interaction (regulation) of the various processes. Based on this functional annotation, a series of *in silico* analyses can be performed to extract biological information from a gene dataset.

Methods

1. Database Implementation (ATGHub)

In order to more accurately identify a list of genes potentially more involved in the autophagy process, this study proposes the creation of a recurrence database, defined as *ATGHub*, based on the presence of various genes in four online resources (databases) related to autophagy. The online resources were selected based on accessibility and update status. For the creation of this dataset, the focus was on protein-coding genes. The resources used were as follows:

Resource	Link	Version	Year	Extracted Genes
AutophagyNet	https://www.autophagynet.org/	v2.1	2022	1228
HADB	https://www.autophagy.lu/	v2	ND	370
HAMdb	http://hamdb.scbdd.com/	v2.4	2018	364
GO:0061919	https://amigo.geneontology.org/amigo	v2.5.17	-	198

These genes were then functionally annotated using the classificatoin of the article “A gene toolbox for monitoring autophagy transcription” (Bordi et al., 2021), hereinafter referred to as “Toolbox 2021”. A flowchart was created to effectively describe the implementation of ATGHub (Figure 1). The Python programming language was used to generate the ATGHub. All scripts used are available in the dedicated GitHub repository (https://github.com/SalvoE276/ATGHub).

In order to select a small set of potentially more relevant genes, the following criterion was adopted: when multiple independent databases or sources annotate the same gene, the probability that the gene is actually involved in a relevant biological process (e.g., disease, pathway, cellular function) increases. In light of this consideration, a final list of genes with the highest recurrence was extracted.

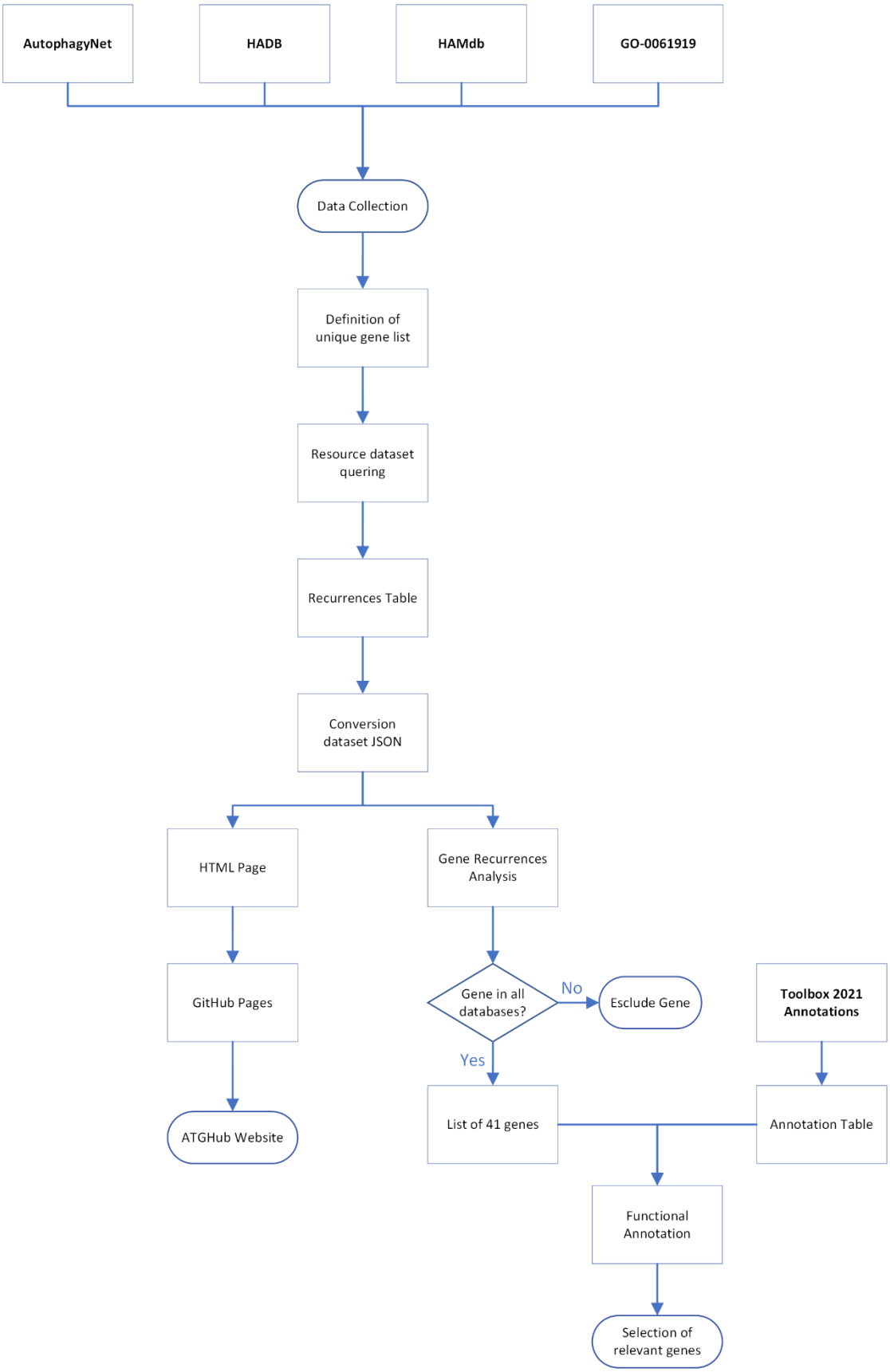

Figure 1. *Flowchart for the implementation of the ATGHub*

1.1. Data Collection

The AutophagyNet data were downloaded in 2 ways from the “Download” section of the website, always in csv format:

1. Download of genes only in “Autophagy core”,
   (file: autophagynet_810933fbcd1232140ef7_csv.csv)
2. Download of genes in “Autophagy core” and in “Direct Regulators (PPI)”,
   (file: autophagynet_05e20118eb7667423678_csv.csv)

The HADB data were downloaded from the “Genes” section in csv format (file: HADB.csv).
The HAMdb data were downloaded from the “Download” section by selecting the csv file “Proteins with autophagy information” (file: protein-role.csv). The data were then filtered to contain only homo sapiens genes (file: protein-role_filter.csv).
The Gene Ontology term was selected to include all autophagic processes. The GO code used was **GO:0061919**, corresponding to *process utilizing autophagic mechanism*. The gene set was downloaded from AmiGO 2 with the following filters:

+ isa_partof_closure: GO:0061919
+ taxon_subset_closure_label: Homo sapiens
- type: protein_complex
+ evidence_subset_closure_label: experimental evidence

Before downloading, the category “Gene/product (bioidentity_label)” was moved to the top to be in the first column of the file (file: GO_0061919.tsv).
All raw data collected were uploaded to the databases_raw_data folder on GitHub.

1.2. Aggregation of Recurrence Data

Before recording the recurrences, a list of genes contained in all databases, without repetitions, and arranged in alphabetical order was generated. Subsequently, for each gene, the Python module of g:Profiler (web version e113_eg59_p19_f6a03c19) (Kolberg et al., 2023) module was used to assign the respective *Entrez ID* and *Ensembl ID* for each gene.
At the end of this process, the databases were queried to verify the presence of the various genes of the list and were annotated with a Boolean system, where “0” indicates absence and “1” indicates presence in the resource. Finally, the dataset was exported in csv format (file: ATGHub.csv), then converted with a Python script to JSON (file: src/dataset_json_convert.py) and finally corrected

manually, at the end of the development of the ATGHub, in its final version viewable in Excel (file: ATGHub_all_dataset_and_annotated.xlsx).

The entire process of aggregating recurrence data was performed using a Python script also available on GitHub (file: ATGHub_aggregator.py).

1.3. Functional Annotation

Functional annotation was performed on the 41 genes present in all databases. After aggregating the data from the Toolbox 2021 Excel file (Bordi et al., 2021), the file was converted to csv, a more accessible format (file: toolbox2021.csv). Automatic annotation was performed using a Python script (file: annotation.py) and exported to csv format (file: ATGHub_annotated.csv). Annotation was then completed manually for not automatically annotated genes.

1.4. Website Front-end Design

To ensure easy access to the database, an ATGHub web page has been set up (https://salvoe276.github.io/ATGHub/). The interface features a search bar for querying by gene symbol. When you type in your search, all results are displayed in a table below. Each column is headed with the name of one of the resources and its associated value is Boolean (1 or 0), indicating the presence or absence in the specific database. The website is designed to be accessed from a desktop interface and is useful for checking the annotation status of a specific gene in the various resources related to autophagy.

The final ATGHub list also includes genes with only one recurrence, which are reported for completeness but should not be considered a priority for biological interpretation unless further evidence is available. Genes with two or more recurrences have greater informational reliability and are preferable for drawing conclusions. However, the relevance of a gene must always be evaluated in relation to the experimental context and the quality of the sources. A limitation of this approach is that some genes may be well known and more frequently cited for historical reasons or experimental bias, while others, although important, may be underrepresented.

Results

1. In Silico Analysis: ATGHub

Given the difficulty of identifying relevant genes for studying, through various techniques, the autophagic process, in this work was generated a dataset of recurrences in various databases dedicated to autophagy with the aim of defining a central hub of ATG genes. This resource was named “ATGHub” and is accessible online for quick and easy consultation. The entire dataset can be downloaded from the GitHub repository. The ATGHub is designed to highlight relevant genes regardless of the experimental model, making it suitable for studying autophagy in multiple biological models.

1.1. Identification of a Central Hub of Genes for Autophagy

The analysis of recurrences led to the identification of a center (defined as *central Hub*) of ATG genes that are potentially relevant for studying the autophagic process. There are 41 genes present in all databases, defining the central Hub (Figure 2). The graph was generated using jvenn software (Bardou et al., 2014).

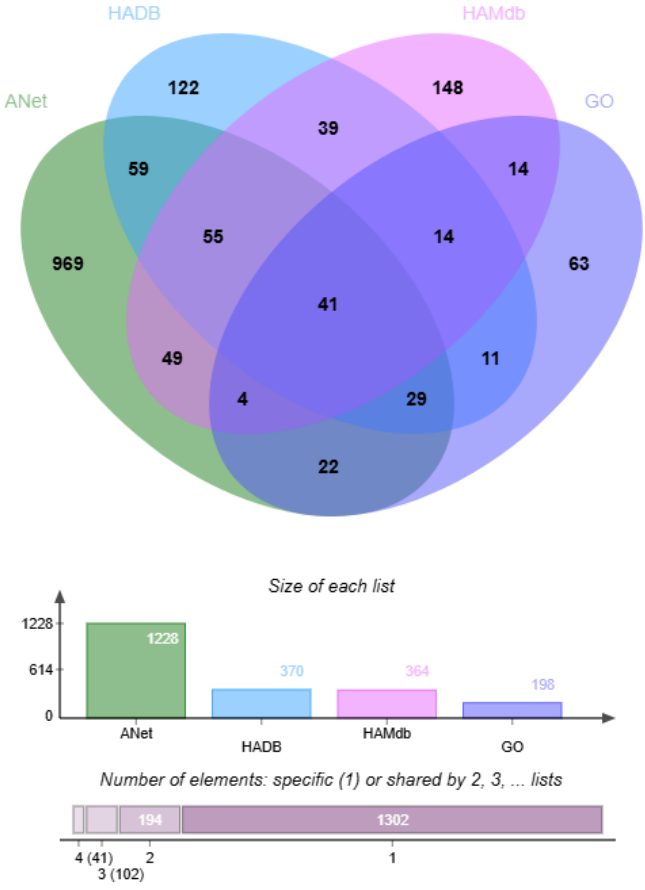

Figure 2. *Distribution of genes in the resources*

The summary table of the 41 genes that form the central Hub, annotated according to Toolbox 2021, is shown below:

Gene Symbol	Entrez ID	Ensembl ID	Category
ATG10	83734	ENSG00000152348	Autophagy core
ATG14	22863	ENSG00000126775	Autophagy core | Docking and fusion
ATG16L1	55054	ENSG00000085978	Autophagy core
ATG2A	23130	ENSG00000110046	Autophagy core
ATG4B	23192	ENSG00000168397	Autophagy core
ATG4C	84938	ENSG00000125703	Autophagy core
ATG5	9474	ENSG00000057663	Autophagy core
ATG7	10533	ENSG00000197548	Autophagy core
ATG9A	79065	ENSG00000198925	Autophagy core
BECN1	8678	ENSG00000126581	Autophagy core | Mitophagy
BNIP3	664	ENSG00000176171	Mitophagy
C9orf72	203228	ENSG00000147894	Autophagy core
CALCOCO2	10241	ENSG00000136436	Autophagy core | Mitophagy
CDKN2A	1029	ENSG00000147889	Other
CHMP4B	128866	ENSG00000101421	Autophagy regulators
HDAC6	10013	ENSG00000094631	Autophagy core | Mitophagy
IRGM	345611	ENSG00000237693	Autophagy core
LAMP2	3920	ENSG00000005893	Lysosome
MAP1LC3A	84557	ENSG00000101460	Autophagy core
MAP1LC3B	81631	ENSG00000140941	Autophagy core
MTOR	2475	ENSG00000198793	mTOR and upstream pathways
NBR1	4077	ENSG00000188554	Autophagy core
NPC1	4864	ENSG00000141458	mTOR and upstream pathways | Lysosome
NRBF2	29982	ENSG00000148572	Autophagy core
OPTN	10133	ENSG00000123240	Autophagy core | Mitophagy
PIK3C3	5289	ENSG00000078142	Autophagy core
PINK1	65018	ENSG00000158828	Autophagy core | Mitophagy
PRKAA1	5562	ENSG00000132356	mTOR and upstream pathways
PRKN	5071	ENSG00000185345	Mitophagy
RPTOR	57521	ENSG00000141564	mTOR and upstream pathways
SQSTM1	8878	ENSG00000161011	Autophagy core | Mitophagy
TECPR1	25851	ENSG00000205356	Autophagy core | Docking and fusion | Lysosome
TMEM74	157753	ENSG00000164841	Autophagy core | Lysosome
TP53	7157	ENSG00000141510	Autophagy regulators
UBQLN1	29979	ENSG00000135018	mTOR and upstream pathways
ULK1	8408	ENSG00000177169	mTOR and upstream pathways | Autophagy core | Mitophagy
ULK2	9706	ENSG00000083290	Autophagy core
UVRAG	7405	ENSG00000198382	Autophagy core
VCP	7415	ENSG00000165280	Mitophagy

WDR45	11152	ENSG00000196998	Autophagy core
ZFYVE1	53349	ENSG00000165861	Autophagy core

Table 1. *List of genes from the ATGHub central Hub with related functional annotations*

The full version of the table with all annotations for the 41 genes can be downloaded from the GitHub repository (file: ATGHub_all_dataset_and_annotated.xlxs). Based on the experimental model and the functional annotation, it is possible to identify the most interesting genes to analyze using various techniques.

1.2. Web Interface for Dataset Querying

To make the ATGHub easily accessible, a web interface has been created at the following URL: https://salvoe276.github.io/ATGHub/.

On the site, you can check the recurrences of a gene in the various autophagy databases by typing the gene symbol in the search bar. As you type, an interactive table is displayed, which also shows values of 0 or 1 to indicate the presence or absence of the gene in the database (Figure 3). In This page there are also QR codes for easy access to resources (Figure 4).

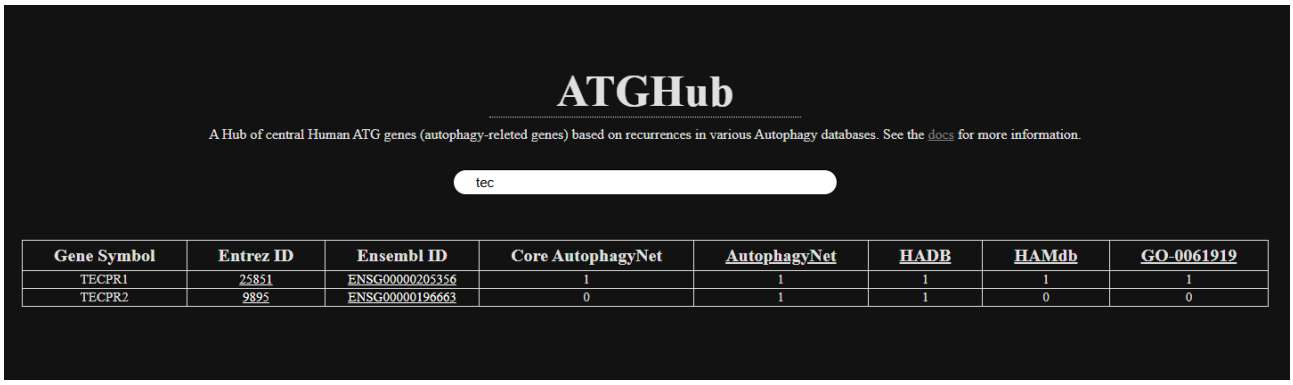

Figure 3*. ATGHub Interface*

Figure 4.
QRcodes of resources

Bibliography

Bardou, P., Mariette, J., Escudié, F., Djemiel, C., & Klopp, C. (2014). Jvenn: An interactive Venn diagram viewer. *BMC Bioinformatics*, *15*(1). https://doi.org/10.1186/1471-2105-15-293

Bordi, M., De Cegli, R., Testa, B., Nixon, R. A., Ballabio, A., & Cecconi, F. (2021). A gene toolbox for monitoring autophagy transcription. *Cell Death and Disease*, *12*(11). https://doi.org/10.1038/s41419-021-04121-9

Csabai, L., Bohár, B., Türei, D., Prabhu, S., Földvári-Nagy, L., Madgwick, M., Fazekas, D., Módos, D., Ölbei, M., Halka, T., Poletti, M., Kornilova, P., Kadlecsik, T., Demeter, A., Szalay-Bekő, M., Kapuy, O., Lenti, K., Vellai, T., Gul, L., & Korcsmáros, T. (2024). AutophagyNet: high-resolution data source for the analysis of autophagy and its regulation. *Autophagy*, *20*(1). https://doi.org/10.1080/15548627.2023.2247737

Kolberg, L., Raudvere, U., Kuzmin, I., Adler, P., Vilo, J., & Peterson, H. (2023). G:Profiler-interoperable web service for functional enrichment analysis and gene identifier mapping (2023 update). *Nucleic Acids Research*, *51*(W1). https://doi.org/10.1093/nar/gkad347

Wang, N. N., Dong, J., Zhang, L., Ouyang, D., Cheng, Y., Chen, A. F., Lu, A. P., & Cao, D. S. (2018). HAMdb: a database of human autophagy modulators with specific pathway and disease information. *Journal of Cheminformatics*, *10*(1). https://doi.org/10.1186/s13321-018-0289-4

Websites

1. **AutophagyNet:** https://www.autophagynet.org/
2. **HAMdb:** http://hamdb.scbdd.com/
3. **HADB:** https://www.autophagy.lu/
4. **jvenn:** https://jvenn.toulouse.inrae.fr/app/index.html
5. **ATGHub GitHub Repository:** https://github.com/SalvoE276/ATGHub

Data Avaliability

All Python scripts are openly avaliable from the dedicated GitHub repository (https://github.com/SalvoE276/ATGHub).

ATGHub Website: https://salvoe276.github.io/ATGHub/