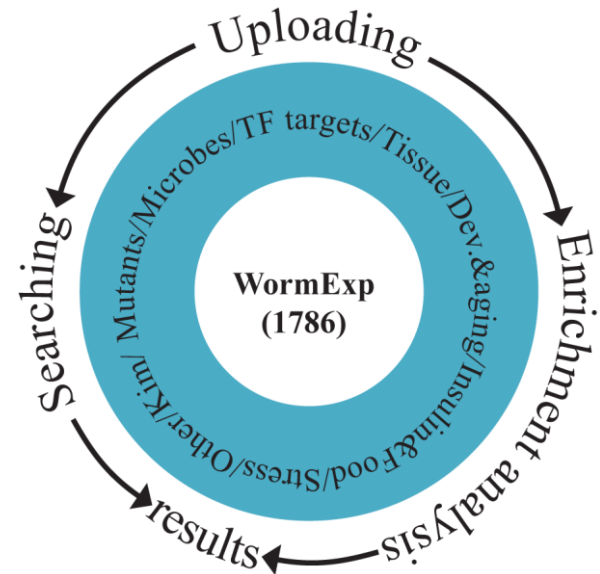## Aims + Workflow

Aims:

WormExp integrates all published expression data for *C. elegans*. It allows performance of taxon-specific enrichment analyses and searches, which may help to identify functions that are taxon-specific and can thus not be captured with GO term or KEGG pathway analyses.

Workflow:

Users start their analysis by uploading a gene list, for example a set of genes, whose expression is induced upon *C. elegans* exposure to a certain condition. The user then has two options: (i) perform an enrichment analysis using either the entire database as reference or selected categories of gene lists (e.g. "mutants"), or (ii) search for overlaps between the uploaded gene set and specific gene lists (e.g. "up *pmk-1* mutant"), selected from the database with the help of keywords (e.g. "pmk-1"). The results of either approach is given in a new window and can be downloaded in table format, including list of the supporting genes.

Text in circle indicates categories of WormExp.

Homepage and Overview

Main sections:
(see red numbers and arrows)
1. **Upload**: uploading list of *C. elegans* gene identifiers (e.g., *clec-4*, Y38E10A.5, WBGene00012583)
2. **Analysis**: Use the uploaded list and select categories to run the enrichment analysis.
3. **Search**: search the database by uploading list of genes or keywords
4. **Download**: download the entire dataset and information
5. **Data categories**: information about the data categories.

Please note that any uploaded data set or any generated result is deleted after 30 min because of memory limitation.

Methods

For the **statistical evaluation during enrichment analysis**, we implemented the method developed for the **program EASE** (Hosack et al. 2003).

Assume a test list with 192 differentially regulated genes, a curated gene set with 2551 genes (e.g., from a previous study), the population size (background of all considered genes) is 27770, and then 34 out of the 192 genes from the test list are present in the curated gene set . Then the 2x2 contigency table for the Fisher Exact test is:

		Background		
		present	absent	total
User case	present	33(34-1)	192-33	192
	absent	2551-33	27770-2551-192+33	27770-192
	total	2551	27770-2551	27770

	Background	
User	33	159
	2518	25060

The one-tailed (greater than 33) p-value is 0.0003292. Detailed information on the statistical approach can be found in the description of the program EASE: Hosack DA, Dennis Jr G, Sherman BT, Lane HC, Lempicki RA: Identifying biological themes within lists of genes with EASE. *Genome Biol* 2003, 4(10):R70.

First step: uploading list of genes

A list of genes (e.g., a list of upregulated genes after cadmium exposure) can be uploaded either as a file (see option "Browse") or by pasting it into the open field. It is possible to upload a query list or a new reference/background gene list.

After upload, statistics for gene list is shown here

Enrichment analysis

An enrichment analysis can be performed with uploaded test list of genes. To start the enrichment analysis, select categories of the curated gene lists or use all of them (none selected – by default). Then click on the "Analyse" button. The results are shown in a separate window or tab.

Results for enrichment analysis

Thresholds: Count 2	**Probability** FDR ▾ 0.1			Refresh the results	will update the result as below!			

Total of Dataset: 1786   ☑ sharedID   ☐ UIDofInput   ☐ UIDofCuratedset   **Download**

Category	Term(select/unselect all)	Counts	ListSize	PopHit	Pop Size	Pvalue	Bonferroni	FDR
Microbes	☑ UP on X. nematophila	15	46	745	27412	1.1e-11	2.0e-08	2.0e-08
Chemicals/stress	☑ UP by Tannic acid 300um	15	46	982	27412	4.3e-10	7.7e-07	3.8e-07
Chemicals/stress	☐ UP dpy-10 mutant	19	46	1934	27412	7.2e-10	1.3e-06	4.2e-07
Mutants	☐ UP by ercc-1 mutant	17	46	1483	27412	1.1e-09	1.9e-06	4.2e-07

The results for the enrichment analysis are shown as a table in a new tab or window. The table includes categories (Category), name of the curated reference gene set (Term), the number of overlapping genes between the test list and the curated gene set (Counts), the total number of genes in the uploaded test list (ListSize), the number of genes in the respective curated gene set (PopHit), background size, that is the total number of genes in the entire considered data base (Pop Size), the inferred p-value before adjustment (Pvalue) and two adjusted p-values (Bonferroni and FDR).

The link underlying the name of the curated gene set leads to more detailed information on the genes contained in this gene set. The user can set thresholds on the minimum number of overlapping genes (Thresholds: Count) or on the probability (Probability; either p-value, bonferroni or FDR).

The results can be downloaded as a table with additional options, including the list of overlapping genes between input and curated list (shared ID; default; options on top of table), or the genes unique for input or for curated gene set (ID of input or ID of curated set) as well as selected curated gene sets (options in second column). Please note that file size may become very large if all three options are chosen.

Based on the uploaded list of genes, the user can search the database for an overlap between the gene list and particular keywords for the curated gene sets. An example for such a keyword is *daf-2*, which will assess all curated gene sets related to this genes, for example transcriptome studies on *daf-2* mutants.


GeneID	Down xbp-1 mutant on daf-2(e1368)	Down xbp-1 mutant on daf-2(e1370)	UP by daf-2;rsks-1 mutant	down by daf-2;rsks-1 mutant
B0511.11	0	0	1	0
lips-10	0	0	0	0
WBGene00012825	0	0	0	0
WBGene00015300	0	0	0	0
WBGene00011483	0	0	0	0
WBGene00023069	0	0	0	0
WBGene00044611	0	0	0	0
WBGene00020066	0	0	1	0
WBGene00044333	0	0	0	0
WBGene00022763	0	0	0	0
WBGene00019436	0	0	1	0
C04E6.4	0	0	0	0
dct-8	0	0	0	1
B0507.10	0	0	1	0
fbxa-224	0	0	0	0

Total of Dataset: 59

Use *daf-2* as an example and click "Search"

The returned results table shows all curated gene sets related to *daf-2*. The rows refer to the genes from the test gene list, whereas columns show the curated gene sets related to *daf-2*. 1 and 0 indicate presence or absence of a gene from the list in the curated gene sets. The results table can be downloaded.

Download datasets
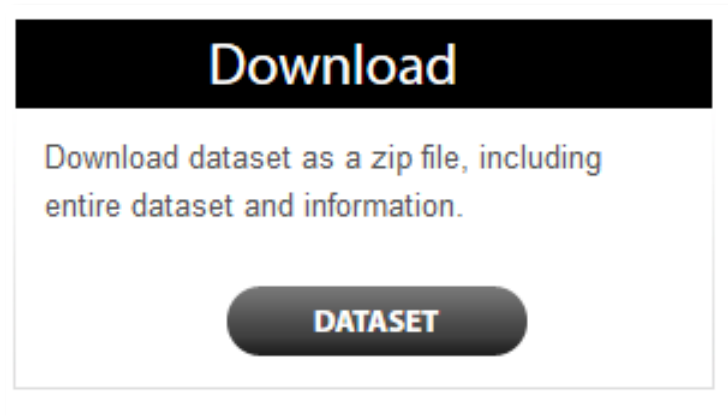
The entire database and further information are available for user download.