

Two new functions in the WormBase Enrichment Suite

David Angeles-Albores¹, Raymond Y. N. Lee¹, Juancarlos Chan¹, and Paul W. Sternberg¹

1. Division of Biology and Biological Engineering, Caltech, Pasadena, CA, 91125, USA

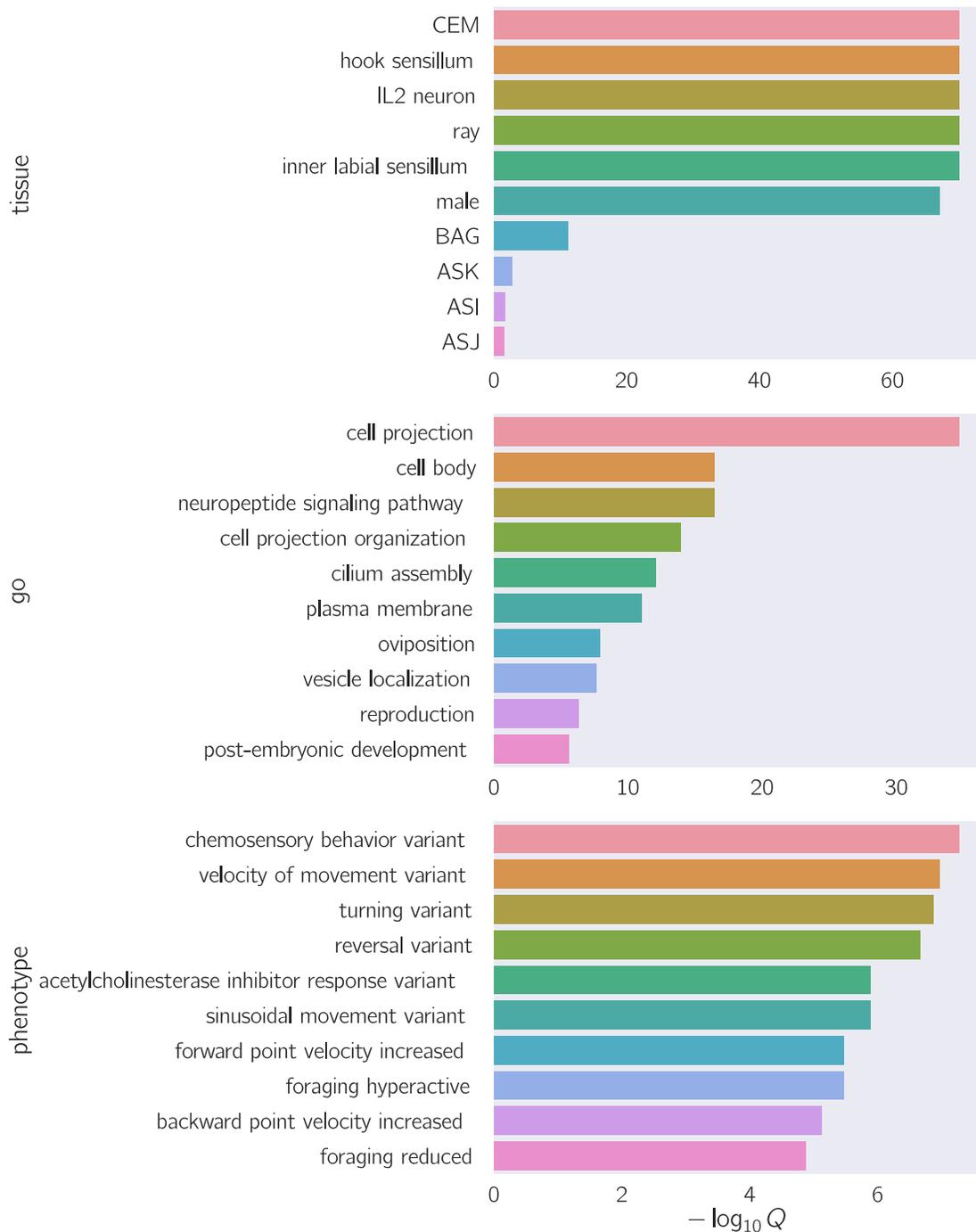


Figure 1. Enrichment results for Tissue, Gene and Phenotype ontologies for genes overexpressed in a ciliary transcriptome (Wang *et al.* 2015).

Description

Genome-wide experiments routinely generate large amounts of data that can be hard to interpret biologically. A common approach to interpreting these results is to employ enrichment analyses of controlled languages, known as ontologies, that describe various biological parameters such as gene molecular or biological function. In *C. elegans*, three distinct ontologies, the Gene Ontology (GO), Anatomy Ontology (AO), and the Worm Phenotype Ontology (WPO) are used to annotate gene function, expression and phenotype, respectively (Ashburner *et al.* 2000; Lee and Sternberg, 2003; Schindelman *et al.* 2011). Previously, we developed software to test datasets for enrichment of anatomical terms, called the Tissue Enrichment Analysis (TEA) tool (Angeles-Albores and Sternberg, 2016). Using the same hypergeometric statistical method, we extend enrichment testing to include WPO and GO, offering a unified approach to enrichment testing in *C. elegans*. The WormBase Enrichment Suite can be accessed via a user-friendly interface at <http://www.wormbase.org/tools/enrichment/tea/tea.cgi>.

To validate the tools, we analyzed a previously published extracellular vesicle (EV)-releasing neuron (EVN) signature gene set derived from dissociated ciliated EV neurons (Wang *et al.* 2015) using WormBase Enrichment Suite based on the WS262 WormBase release. TEA correctly identified the CEM, hook sensillum and IL2 neuron as enriched tissues. The top phenotype associated with the EVN signature was chemosensory behavior. Gene Ontology enrichment analysis showed that cell projection and cell body were the most enriched cellular components in this gene set, followed by the biological processes neuropeptide signaling pathway and vesicle localization further down. The tutorial script used to generate the figure above can be viewed at:

<https://github.com/dangeles/TissueEnrichmentAnalysis/blob/master/tutorial/Tutorial.ipynb>

The addition of Gene Enrichment Analysis (GEA) and Phenotype Enrichment Analysis (PEA) to WormBase marks an important step towards a unified set of analyses that can help researchers to understand genomic datasets. These enrichment analyses will allow the community to fully benefit from the data curation ongoing at WormBase.

Method

Using the methods described in Angeles-Albores *et al.*, we generated ontology dictionaries using the Anatomy, Phenotype and Gene Ontology annotations for *C. elegans*. The dictionary similarity parameter was set to 95% for all ontologies. The annotation per term minimum was set to 33 annotations for the AO, a 50 annotations for the WPO, and 33 annotations for GO. Terms within the dictionary are tested using a hypergeometric probability test and corrected using the Benjamini-Hochberg step-up algorithm. In WS262, there are 1320 anatomy terms, 1117 phenotypes, and 3025 GO terms that have at least 11 genes annotated to them. The dictionaries are freely accessible using the Python version of the Suite, which can be installed using the pip tool for Python libraries:

```
pip install tissue_enrichment_analysis
```

The dictionary can then be automatically downloaded by importing the enrichment analysis library in a Python script by writing:

```
import tissue_enrichment_analysis as ea
```

The dictionaries can then be downloaded by typing:

```
ea.fetch_dictionary(dict)
```

into Python, where `dict` is the string `tissue`, `phenotype` or `go` to specify which dictionary to download. If the function does not receive an argument, the dictionary corresponding to the AO is downloaded by default. See the tutorial above for an example implementation.

References

Angeles-Albores D, N. Lee RY, Chan J, Sternberg PW: Tissue enrichment analysis for *C. elegans* genomics. *BMC Bioinformatics* 2016 Sep; 17(1): 366 PubMed PMID 27618863

03/27/2018 – Open Access

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G, Consortium GO: Gene Ontology: tool for the unification of biology. *Nat Genet* 2000 May; 25(1): 25–29 PubMed PMID 10802651

Lee RYN, Sternberg PW: Building a cell and anatomy ontology of *Caenorhabditis elegans*. *Comparative and Functional Genomics* 2003 Feb;4(1):121–126 PubMed PMID 18629098

Schindelman G, Fernandes JS, Bastiani CA, Yook K, Sternberg PW: Worm Phenotype Ontology: integrating phenotype data within and beyond the *C. elegans* community. *BMC Bioinformatics* 2011 Jan; 24; 12: 32 PubMed PMID 21261995

Wang J, Kaletsky R, Silva M, Williams A, Haas LA, Androwski RJ, Landis JN, Patrick C, Rashid A, Santiago-Martinez D, Gravato-Nobre M, Hodgkin J, Hall DH, Murphy CT, Barr MM: Cell-Specific Transcriptional Profiling of Ciliated Sensory Neurons Reveals Regulators of Behavior and Extracellular Vesicle Biogenesis. *Curr Biol* 2015 Dec; 25: 3232–3238 PubMed PMID 26687621

Funding This work was supported by the NIH grant U41 HG002223.

Reviewed by Paola Roncaglia

Received 02/07/2018, **Accepted** 02/28/2018. **Available** starting WormBase release WS266, **Published Online** 03/27/2018.

Copyright © 2018. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Citation Angeles-Albores, D; Lee, RYN; Chan, J; Sternberg, PW (2018): Two new functions in the WormBase Enrichment Suite. Micropublication: biology. Dataset. <https://doi.org/10.17912/W25Q2N>