

Folder Structure

In the SI, we provide the entire output of our tests for the interested reader. The folder is organized as follows:

HGTXX_any_Results

The folders titled HGTXX_any_Results contain the enrichment results of all the 30 golden sets tested with a dictionary with cutoff XX, similarity threshold = 0.95 and thresholding method ‘any’. The folders contain two kinds of files, ‘pdf’ and ‘csv’ files. The pdf files contain the bar charts for each analysis, whereas the csv files contain the complete output for a particular geneset. There is also a file called ‘empty.txt’ that contains the names of any gene sets where no terms were enriched.

Files are titled according to a common schema. For example, the file `WBPaper00013489_Ray_Enriched_WBbt_0006941_25.csv` refers to the **WormBase Paper 0013489**, which should be ‘**Ray Enriched**’, and specifically should be enriched in term ‘**WBbt:0006941**’. The gene set that was used for this analysis is contained in the SI folder named ‘**golden gene sets**’ and is contained in a homologically named csv file.

Engelmann

Contains all graphs pertaining to the data from Engelmann 2011.