

Inputs

Algorithm

Obo file, or
local or remote
GO term
database, or
pre-cached
ontology

1. Parse GO ontology

Parse GO term hierarchy for parent-child relationships, obsolete terms and preferred synonymous terms.

NCBI gene2go
table, or
Affymetrix or
Agilent array
annotation file,
or local or
remote GO
annotation
database

2. Obtain gene associations, filter and map IDs

Obtain gene to GO term associations.

Filter by evidence code.

Select gene ID or symbol as required.

Optionally map annotated gene identifiers to user supplied alternative identifiers.

Map annotated GO term identifiers to preferred synonymous identifiers in the supplied GO hierarchy.

Warn the user about obsolete or missing GO term identifiers.

3. Annotate genes with higher level term associations

Recursively propagate gene associations towards the root of the DAG, recording each term's distance from the root.

4. Size filtering, duplicate removal and output

Make a gene set for each GO term containing all the genes annotated with that term after propagation.

Filter sets by max and min set size.

Prune identical (by content) sets.

Output in .gmt or .gmx format.