Ontologies

What is an ontology?

An ontology is a hierachical, organized dictionary of terms. It provides a standard set of labels for researchers to apply to genes, and its structure allows for reasoning about terms.

What data does each ontology provide?

GREAT uses numerous ontologies covering a wide range of topics, which allows you to examine many aspects of your input set. Currently, GREAT includes the Gene Ontology (GO) and other ontologies covering phenotypes and human disease, pathways, gene expression, regulatory motifs, and gene families.

What statistics are given for each ontology listed below?

We provide three statistics for each ontology listed below in the footer of the GREAT output tables:

- **Terms**: The number of terms in the ontology that GREAT tests
    - This number may be smaller than the total number of terms defined by the source ontology, for two reasons. If an ontology term is not associated with any genes in the gene set used by GREAT, we do not test it since by definition that term will never be enriched. Additionally, redundant terms are omitted: if term **P** is a parent of term **C** in the ontology, and the genes associated with terms **P** and **C** are identical, we do not test **P** since it will give identical enrichment scores as **C** and is a less specific term.
- **Genes**: The number of genes in the GREAT gene set that are annotated with one or more of the terms above
    - Term <--> gene associations: The number of unique associations between terms and genes used by GREAT
- **The term <--> gene association**: number is calculated after propagating associations from child terms to parent terms (e.g., if gene **g** is annotated with ontology term **C**, and **C** is a child of term **P**, then **g** implicitly receives the annotation **P**) and omitting associations to redundant terms (see above).

Ontologies for human

Gene Ontology (GO)

- GO Molecular Function
- GO Biological Process
- GO Cellular Component

Phenotype Data and Human Disease

- Mouse Phenotype - data about mouse genotype - phenotype associations mapped to human genes
- Human Phenotype - data of human phenotypes mapped to human genes
- Disease Ontology - data linking human diseases to associated genes
- MSigDB Cancer Neighborhood - gene sets defined by correlated expression profiles with cancer-associated genes
- Placenta Disorders - data on genes associated with pre-eclampsia and preterm birth

Pathway Data

- PANTHER Pathway - primarily signaling pathways
- Pathway Commons - a collection of pathways from multiple sources
- BioCyc Pathway - metabolic pathways
- MSigDB Pathway - gene sets from pathway databases

Gene Expression

- MGI Expression Detected - genes detected as expressed in data with a focus on gene expression during mouse development
- MSigDB Perturbation - gene sets that change their expression after genetic and chemical perturbations

Regulatory Motifs

- Transcription Factor Targets - genes identified by a transcription factor ChIP-chip experiment
- MSigDB Predicted Promoter Motifs - gene sets that share a transcription factor binding motif
- MSigDB miRNA Motifs - gene sets that share a 3'-UTR microRNA binding motif
- miRNA Targets - genes that are downregulated in a microRNA overexpression experiment

Gene Families

- InterPro - protein domains, families and functional sites
- TreeFam - gene families of paralogs
- HGNC Gene Families - gene sets based on sequence similarity, data from the literature, and manual curation

Ontologies for mouse

Gene Ontology (GO)

- GO Molecular Function
- GO Biological Process
- GO Cellular Component

Phenotype Data

- Mouse Phenotype - data about mouse genotype - phenotype associations primarily obtained by curators from the literature
- Human Phenotype - data of human phenotypes mapped to human genes
- Disease Ontology - data linking human diseases to associated genes

Pathway Data

- PANTHER Pathway - primarily signaling pathways
- Pathway Commons - a collection of pathways from multiple sources
- BioCyc Pathway - metabolic pathways
- MSigDB Pathway - gene sets from pathway databases

Gene Expression

- MGI Expression Detected - genes detected as expressed in data with a focus on gene expression during mouse development
- MSigDB Perturbation - gene sets that change their expression after genetic and chemical perturbations

Regulatory Motifs

- Transcription Factor Targets - genes identified by a transcription factor ChIP-chip experiment
- MSigDB Predicted Promoter Motifs - gene sets that share a transcription factor binding motif
- MSigDB miRNA Motifs - gene sets that share a 3'-UTR microRNA binding motif
- miRNA Targets - genes that are downregulated in a microRNA overexpression experiment

Gene Families

- InterPro - protein domains, families and functional sites
- TreeFam - gene families of paralogs

Ontologies for zebrafish

Gene Ontology (GO)

- GO Molecular Function
- GO Biological Process

- GO Cellular Component

Phenotype Data

- Zebrafish Phenotype - data about zebrafish genotype - phenotype associations primarily obtained by curators from the literature

Pathway Data

- Wiki Pathways - zebrafish pathway data

Gene Expression

- Zebrafish Wildtype Expression - genes detected as expressed in wildtype zebrafish during development

Gene Families

- InterPro - protein domains, families and functional sites
- TreeFam - gene families of paralogs

Previous GREAT version ontology statistics

- GREAT v1.8 Ontologies
- GREAT v1.2 Ontologies

Can I use other ontologies?

Currently, GREAT only supports the listed ontologies. Feel free to contact us and suggest additional ontologies you'd like to see us add.