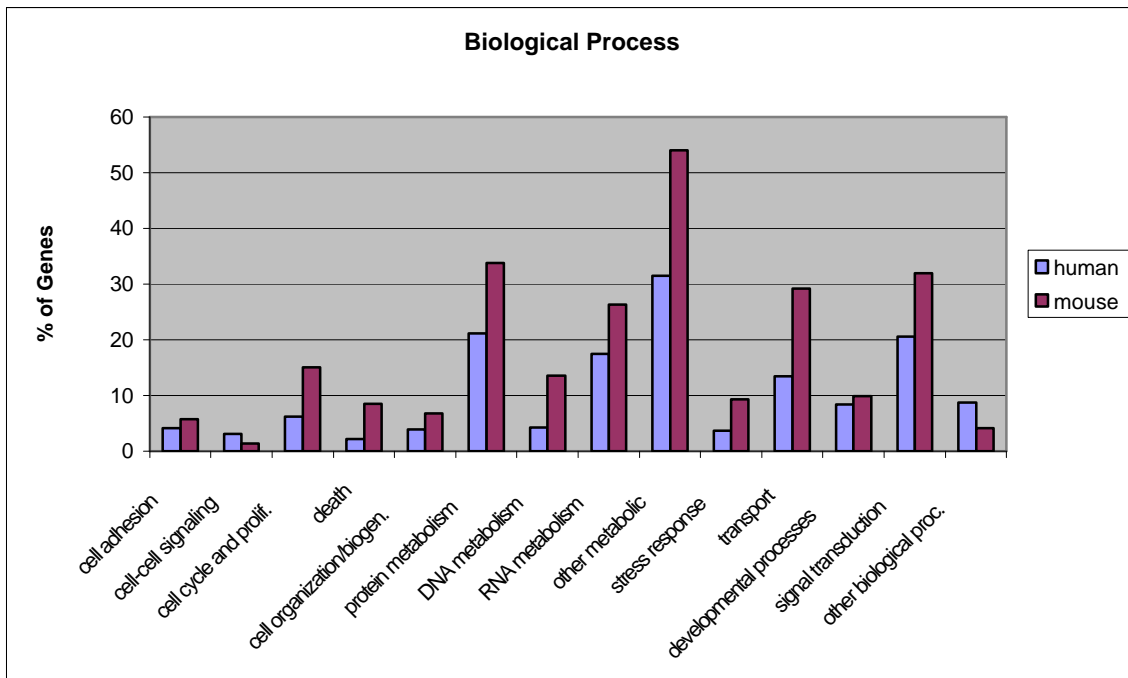
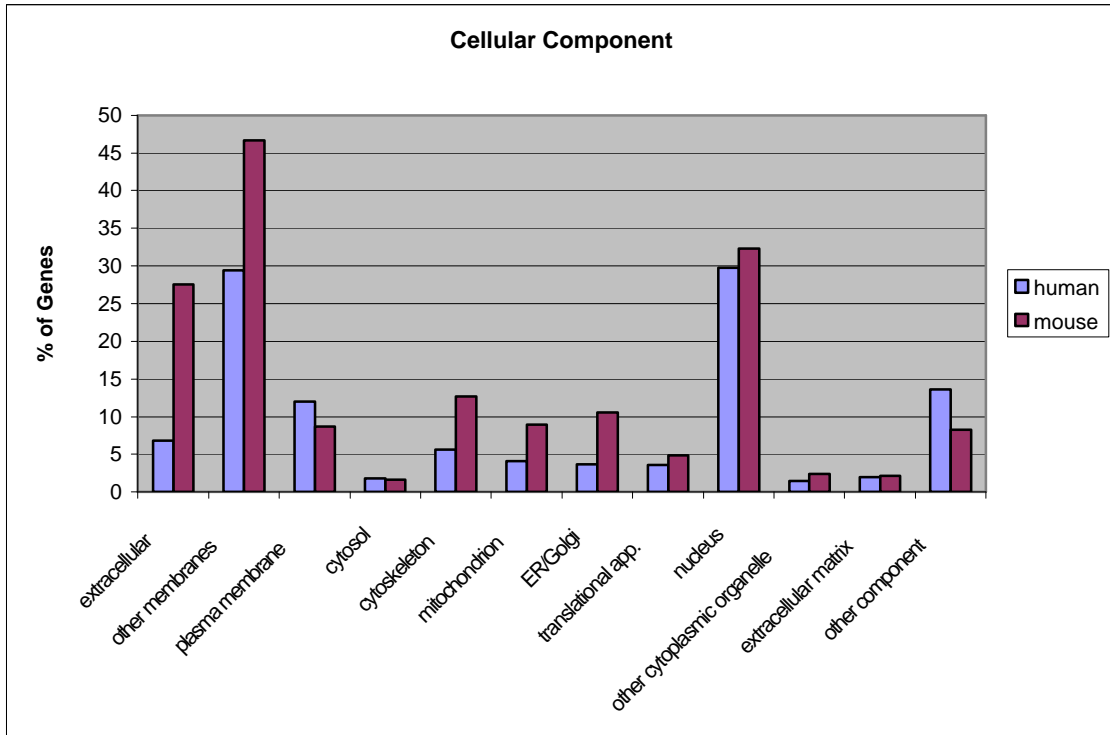


Supplementary Information 16



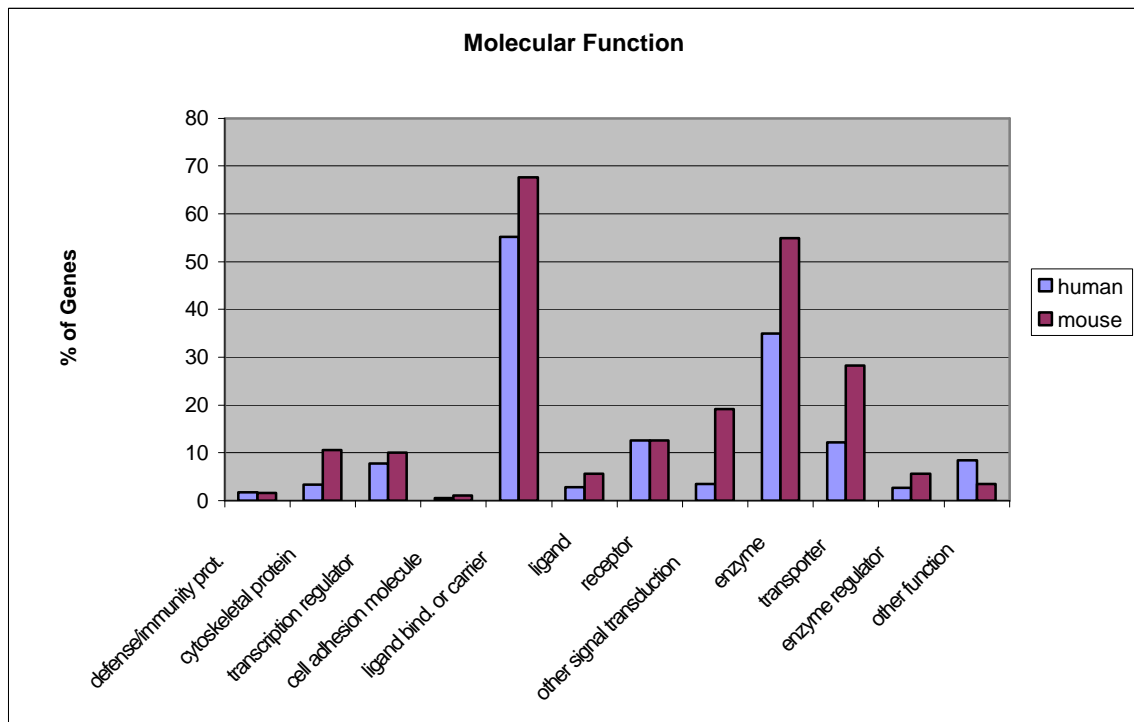
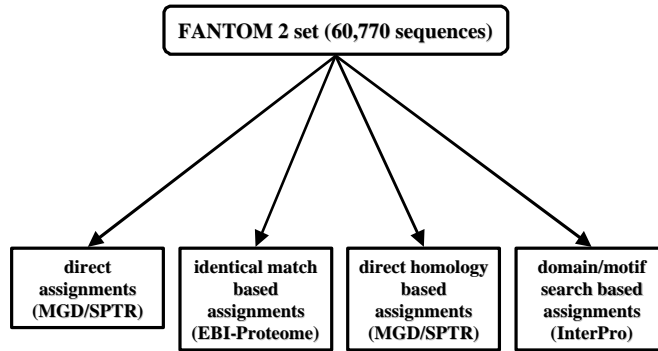


Figure GO

Binning of GO assignments into categories represented by the GO vocabularies. Each chart shows the percentage of transcription units that fall into a particular bin for each of the three GO vocabularies. The strategies are shown in supplemental data and the July release of the GO vocabularies were used for the binning. The left bar of each pair represents data taken from the ENSEMBL annotation of human gene products (http://www.geneontology.org/cgi-bin/GO/downloadGOGA.pl/gene_association.goa_human) and the right bar represents data from the mouse RPTS.

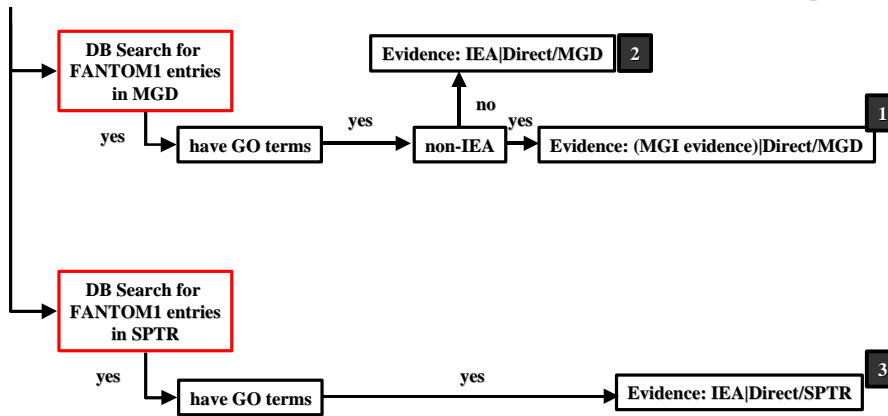
Computational Assignment of GO terms

Automated GO assignment pipeline overview

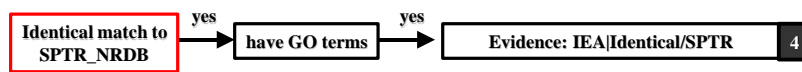


1: Direct Assignments

NOTE: smaller number means higher priority in case that same GO terms are assigned



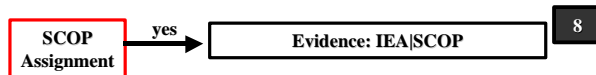
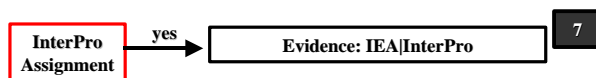
2: Identical Match-based Assignments



3: Direct Homology-based Assignments



4: Domain/Motif Assignments



CODE	Definition
Direct/MGD	Directly assigned in MGI
Direct/SPTR	Directly assigned in SPTR
Identical/SPTR	Assigned from identically matched SPTR protein sequences
BLASTN/MGD	Assigned from MGD DNA sequences matched with BLASTN searches
FASTY/SPTR	Assigned from SPTR protein sequences matched with FASTY searches
InterPro	Assigned from InterPro domains
SCOP	Assigned from SCOP/SuperFamily domains
EXP	Assigned after analysis by expert curation

GO Binning strategies

In all cases when a term was searched the search included annotations to all other terms that are parts of or types of that term. In some cases a particular term was EXCLUDED from a search. The OR operator indicates the union of the gene sets from searches.

The NOT operator indicates that the genes annotated to that term were subtracted from the rest of the search. The searches are as follows:

Molecular Function bins

- 1.) defense/immunity protein: defense/immunity protein
- 2.) cytoskeletal protein: cytoskeletal regulator OR motor OR structural constituent of cytoskeleton OR structural constituent of eye lens OR structural constituent of muscle OR cytoskeletal binding protein
- 3.) transcription regulator: transcription regulator
- 4.) cell adhesion molecule: cell adhesion molecule
- 5.) ligand binding or carrier: ligand binding or carrier
- 6.) ligand: ligand

- 7.) receptor: receptor
- 8.) other signal transduction molecule: signal transducer EXCLUDING (ligand OR receptor)
- 9.) enzyme: enzyme
- 10.) transporter: transporter
- 11.) enzyme regulator: enzyme regulator
- 12.) other molecular function: NOT (1-11)

Biological Process Bins

- 1.) cell adhesion: cell adhesion
- 2.) cell-cell signaling: cell-cell signaling
- 3.) cell cycle and proliferation: cell cycle OR cell proliferation
- 4.) death: death
- 5.) cell organization and biogenesis: cell organization and biogenesis
- 6.) protein metabolism: protein metabolism
- 7.) DNA metabolism: DNA metabolism
- 8.) RNA metabolism: RNA metabolism OR transcription
- 9.) other metabolic processes: metabolism EXCLUDING (DNA metabolism OR RNA metabolism)
- 10.) stress response: stress response
- 11.) transport: transport
- 12.) developmental processes: developmental processes
- 13.) signal transduction: signal transduction
- 14.) other biological processes: NOT (1-12)

Cellular Component Bins

- 1.) non-structural extracellular: extracellular EXCLUDING extracellular matrix
- 2.) extracellular matrix: extracellular matrix
- 3.) plasma membrane: plasma membrane
- 4.) other membranes: (membrane EXCLUDING plasma membrane) OR (membrane fraction NOT plasma membrane)
- 5.) cytosol: cytosol OR (sarcoplasm EXCLUDING (sarcoplasmic reticulum OR junctional membrane complex))

- 6.) cytoskeleton: cytoskeleton OR microtubule organizing center OR spindle OR muscle fiber OR cilia OR flagellum (sensu Eukarya)
- 7.) mitochondrion: mitochondrion
- 8.) ER/Golgi: endoplasmic reticulum OR ER-Golgi intermediate compartment OR Golgi apparatus OR transport vesicle OR Golgi vesicle
- 9.) translational apparatus: eukaryotic 43S pre-initiation complex OR eukaryotic 48S initiation complex OR eukaryotic translation initiation factor 2B complex OR eukaryotic translation initiation factor 4F complex OR nascent polypeptide-associated complex OR signal sequence receptor complex OR ribosome
- 10.) nucleus: nucleus
- 11.) other cytoplasmic organelle: acidocalcisome OR cytoplasmic exosome OR endosome OR glyoxysome OR lysosome OR peroxisome OR vacuole
- 12.) other cell component: cellular component NOT (1-11)