Ontological Visualization of Protein-Protein Interaction

Harold J. Drabkin
Christopher Hollnenbec
David P. Hill
Judith A. Blake

Outline

- Proteomics
- Protein-protein interaction
- The Gene Ontology database
- Specific Vs Network Interactions
- Assigning protein function

Proteomics

- Genome → Proteome
- Genome is merely a parts list
- Identifying protein-protein interactions is one of the most challenging problems of the post-genomic era

Protein-Protein Interaction

- Aim is to understand the mechanism by which a set of proteins communicate and collaborate toward a common cellular function

- This will allow a broader outlook and a more comprehensive understanding of the biological processes occurring in cells

Protein-Protein Interaction

- Understanding protein interactions requires two steps
  - Identifying proteins that interact (experimentally and computationally)
  - Determining the significance of the interaction network

Protein-Protein Interaction

- There is enormous amount of specific protein-protein interaction data available (by studies focusing on a few interacting proteins)

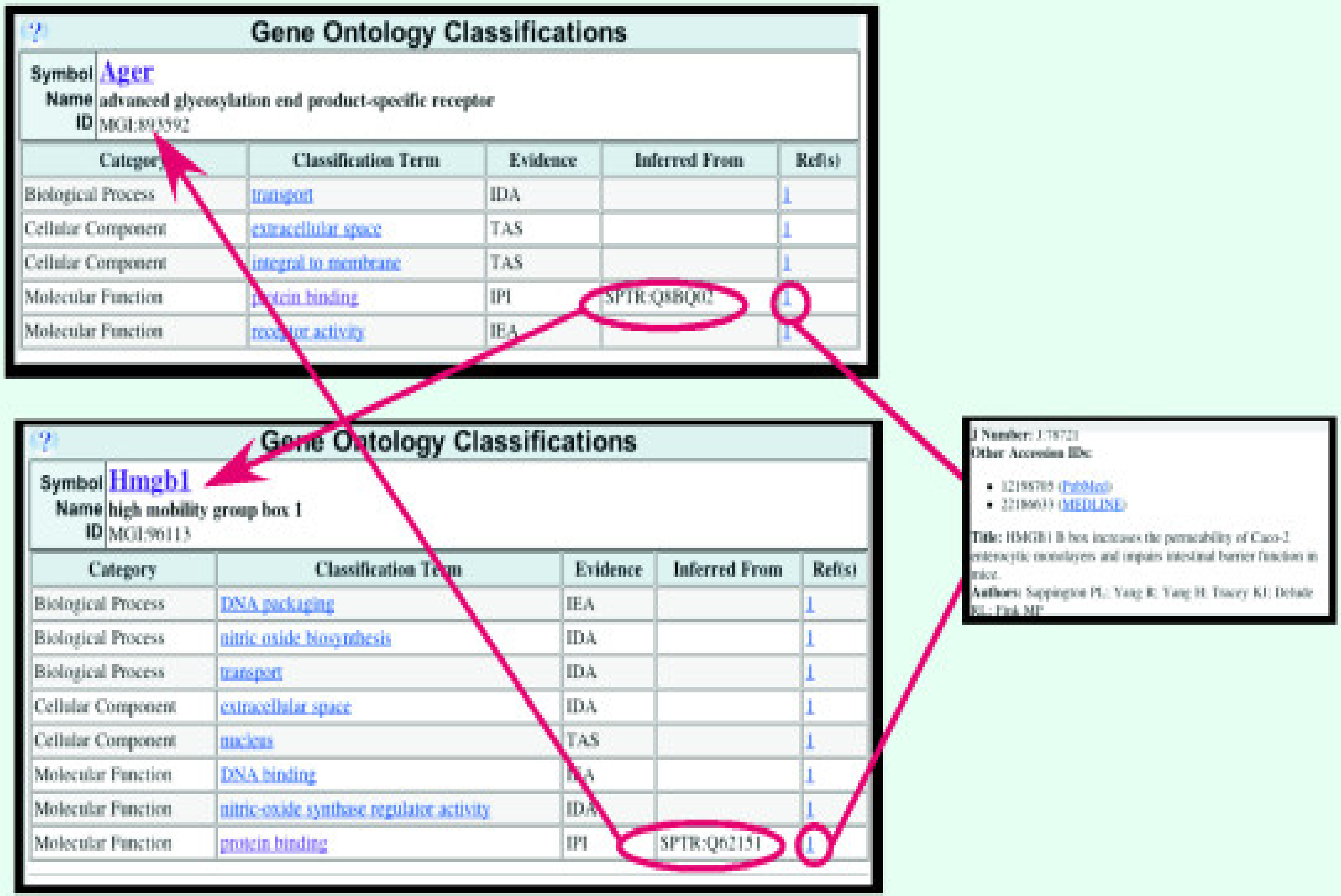
- There is a need to extract data from different studies and accumulate all data into one databank

The Gene Ontology (GO) Database

- Provides a description of genes and how gene products behave in a cellular context
- Ontology (vocabulary set): cellular component, biological process, and molecular function
- Annotation is the characterization of gene products using terms from the ontology

The Gene Ontology (GO) Database

- Information encoded in GO annotations was used to develop protein interaction networks for mouse
- A gene product can be annotated to “protein binding” using the IPI (inferred from physical interaction)
- Ex: Ager protein binds to Q8BQ02



Gene Ontology Classifications

Symbol Ager
Name advanced glycosylation end product-specific receptor
ID MGI:893592

Category	Classification Term	Evidence	Inferred From	Ref(s)
Biological Process	transport	IDA		1
Cellular Component	extracellular space	TAS		1
Cellular Component	integral to membrane	TAS		1
Molecular Function	protein binding	IPI	SPTR:Q8BQ02	1
Molecular Function	receptor activity	IEA		1

Gene Ontology Classifications

Symbol Hmgb1
Name high mobility group box 1
ID MGI:96113

Category	Classification Term	Evidence	Inferred From	Ref(s)
Biological Process	DNA packaging	IEA		1
Biological Process	nitric oxide biosynthesis	IDA		1
Biological Process	transport	IDA		1
Cellular Component	extracellular space	IDA		1
Cellular Component	nucleus	TAS		1
Molecular Function	DNA binding	IEA		1
Molecular Function	nitric-oxide synthase regulator activity	IDA		1
Molecular Function	protein binding	IPI	SPTR:Q62151	1

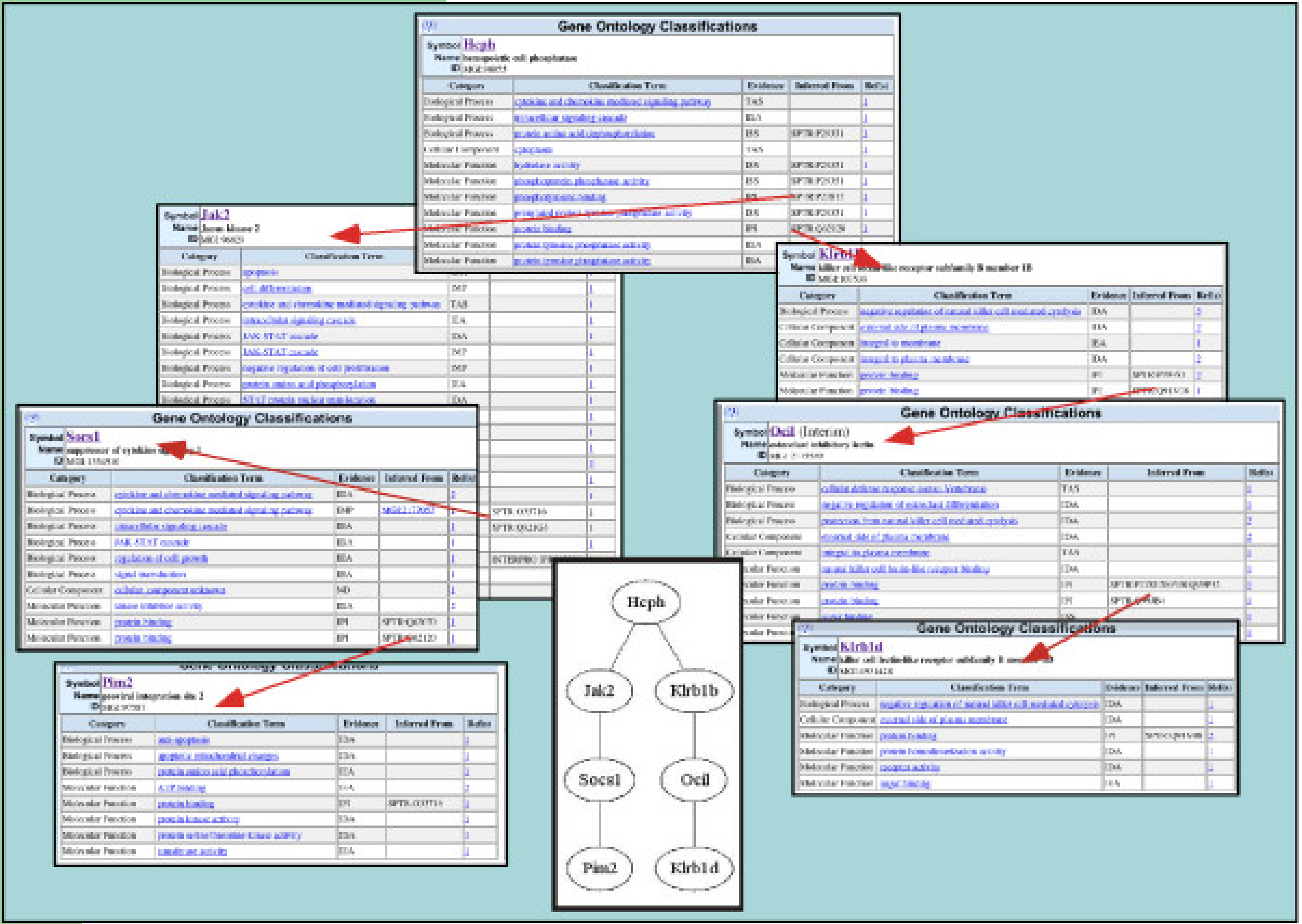
J Number: J:78721
Other Accession IDs:

- 12198705 (PubMed)
- 22180633 (MEDLINE)

Title: HMGB1 B box increases the permeability of Caco-2 enterocytic monolayers and impairs intestinal barrier function in mice.
Authors: Sappington PL; Yang R; Yang H; Tracey KJ; Delude RL; Fink MP

Specific Interaction Vs Networks

- Ager with Q8BQ02
- Protein could have further annotations
- Ex. Of simple network
  - Hcph (hemopoetic cell phosphate) binds to jak 2 and klrb1b. Jak2 not only binds to hcph but also SOCS1 which in turn binds with PIM2

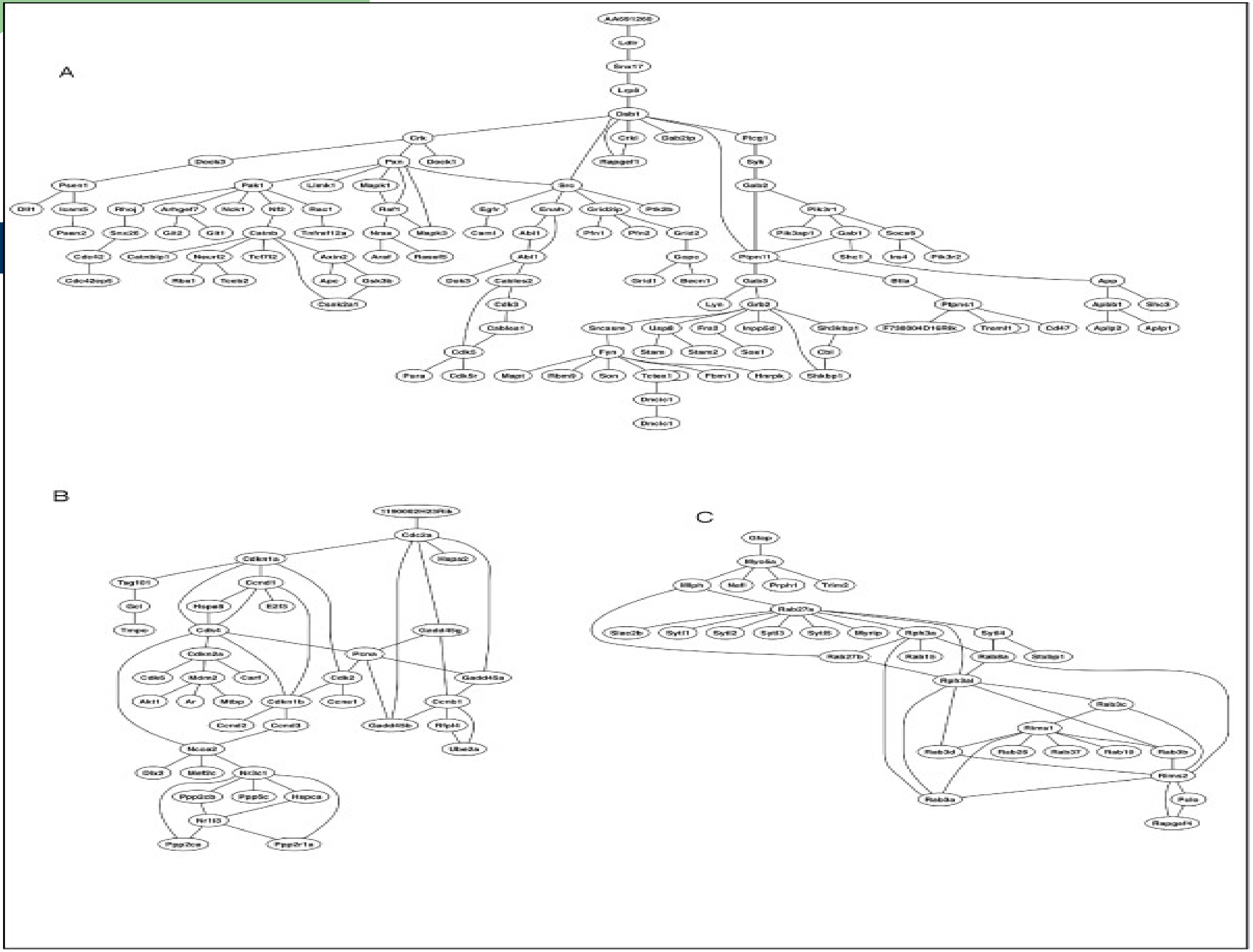

Gene Ontology Classifications
Hcph
Jak2
Klrb1b
Gene Ontology Classifications
Socs1
Gene Ontology Classifications
Ocil (Interim)
Gene Ontology Classifications
Klrb1d
Pim2
Category
Classification Term
Evidence
Inferred From
Ref(s)
Hcph
Jak2
Klrb1b
Socs1
Ocil
Pim2
Klrb1d

Proteins with unknown function

- A protein is known to interact with another protein but there is no information known about the biological role of this protein
- Use the GO annotation to determine the function of the interacting proteins and use this to assign a function to the unknown protein

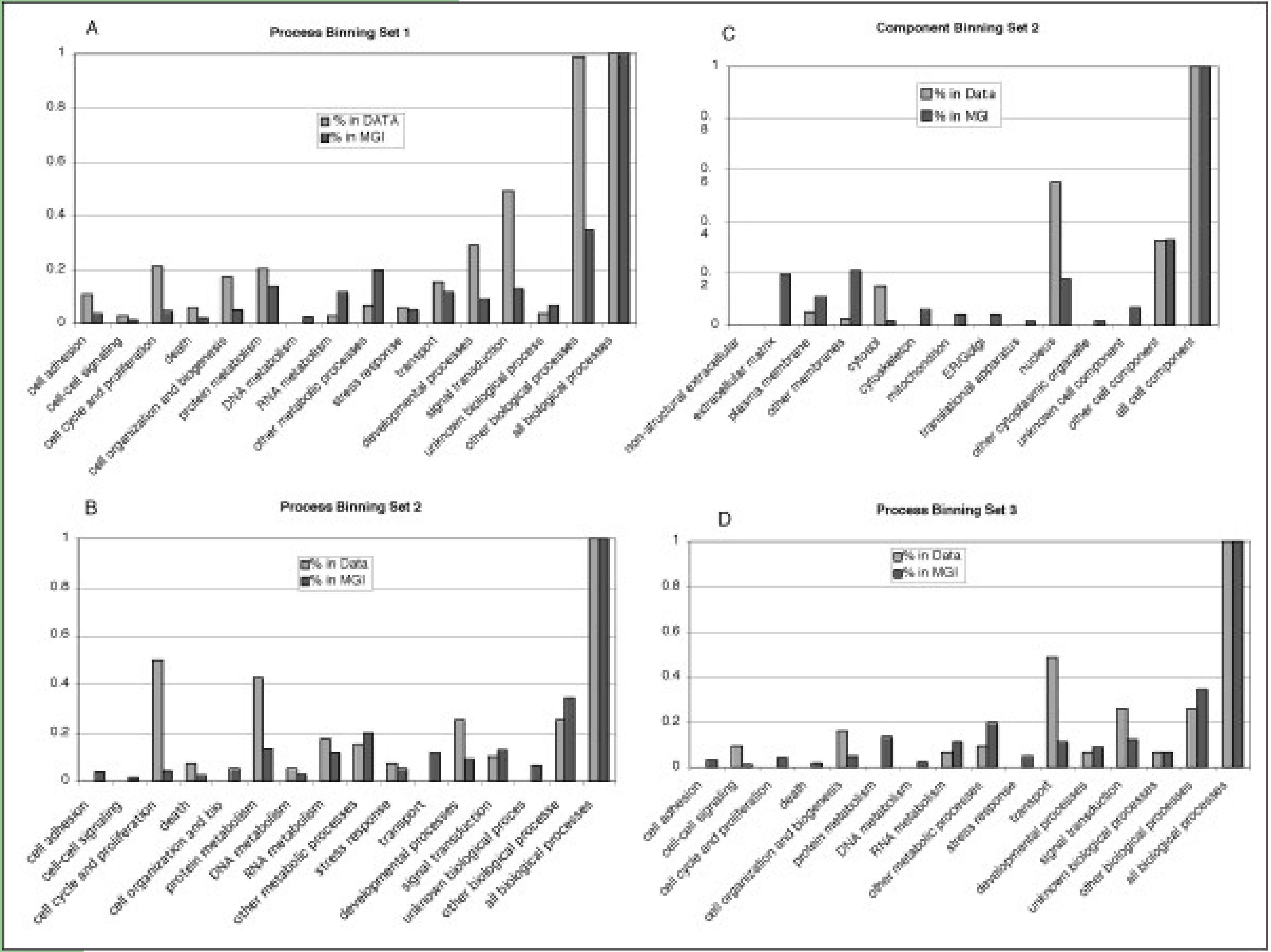
GO Tools

- GO Term Finder and its graphical interface Vlad searches for shared GO Terms and organizes information about the roles of proteins


A
Ldlr
Snx17
Lrp8
Dab1
Crk
Crkl
Gab2ip
Plcg1
Dock3
Pxn
Dock1
Rapgef1
Syk
Psen1
Pak1
Limk1
Mapk1
Src
Gab2
Dll1
Icam5
Rhoj
Arhgef7
Nck1
Nf2
Rac1
Raf1
Egfr
Enah
Grid2ip
Ptk2b
Pik3r1
Psen2
Snx26
Git2
Git1
Ctnnb
Tnfrsf12a
Nras
Mapk3
Caml
Abl1
Pfn1
Pfn2
Grid2
Pik3ap1
Gab1
Socs5
Cdc42
Ctnnbip1
Neurl2
Tcf7l2
Axin2
Araf
Rassf5
Gopc
Ptpn11
Shc1
Irs4
Pik3r2
Cdc42ep5
Rbe1
Tcea2
Apc
Gsk3b
Cdk3
Cables2
Grid1
Becn1
Gab3
App
Csnk2a1
Lyn
Grb2
Ptpns1
Apbb1
Shc3
Cdk5
Cables1
Sncaip
Usp8
Frs3
Inpp5d
Shkbp1
F730004D16Rik
Trem1
Cd47
Aplp2
Aplp1
Fyn
Stam
Stam2
Sos1
Cbl
Pura
Cdk5r
Mapt
Rbm9
Son
Tctex1
Fbn1
Hnrpk
Dnclc1
B
1190002H23Rik
Cdc2a
Cdkn1a
Hspa2
Tsg101
Ccnd1
Gcl
Hspa8
E2f3
Timp
Cdk4
Gadd45g
Cdkn2a
Pcna
Cdk6
Mdm2
Cdk2
Gadd45a
Akt1
Ar
Mtbp
Cdkn1b
Ccne1
Ccnb1
Ccnd2
Ccnd3
Gadd45b
Rtp4
Ube3a
Ncoa2
Nr3c1
Ppp2cb
Ppp5c
Hspca
Nr1i3
Ppp2ca
Ppp2r1a
C
Myo5a
Mlph
Prph1
Trim2
Rab27a
Slac2b
Sytl1
Sytl2
Sytl3
Sytl5
Myrip
Rph3a
Syt4
Rab27b
Rab15
Rab8a
Rph3al
Rab3c
Rims1
Rab3d
Rab26
Rab37
Rab10
Rab3b
Rims2
Rab3a
Pclo
Rapgef4

GO Tools

- GO Slim Viewer provides a summary of the GO categories for a dataset based on specific annotations


A
Process Binning Set 1
% in DATA
% in MGI
1
0.8
0.6
0.4
0.2
0
cell adhesion
cell-cell signaling
cell cycle and proliferation
death
cell organization and biogenesis
protein metabolism
DNA metabolism
RNA metabolism
other metabolic processes
stress response
transport
developmental processes
signal transduction
unknown biological process
other biological processes
all biological processes
C
Component Binning Set 2
% in Data
% in MGI
1
0.8
0.6
0.4
0.2
0
non-structural extracellular
extracellular matrix
plasma membrane
other membranes
cytosol
cytoskeleton
mitochondrion
ER/Golgi
translational apparatus
nucleus
other cytoplasmic organelle
unknown cell component
other cell component
all cell component
B
Process Binning Set 2
% in Data
% in MGI
1
0.8
0.6
0.4
0.2
0
cell adhesion
cell-cell signaling
cell cycle and proliferation
death
cell organization and bio
protein metabolism
DNA metabolism
RNA metabolism
other metabolic processes
stress response
transport
developmental processes
signal transduction
unknown biological proces
other biological process
all biological processes
D
Process Binning Set 3
% in Data
% in MGI
1
0.8
0.6
0.4
0.2
0
cell adhesion
cell-cell signaling
cell cycle and proliferation
death
cell organization and biogenesis
protein metabolism
DNA metabolism
RNA metabolism
other metabolic processes
stress response
transport
developmental processes
signal transduction
unknown biological processes
other biological processes
all biological processes

Conclusion

- Gene annotations for protein binding interactions are based on published literature supported by experimental evidence
- GO annotations can be used to visualize and analyze protein-protein interaction networks