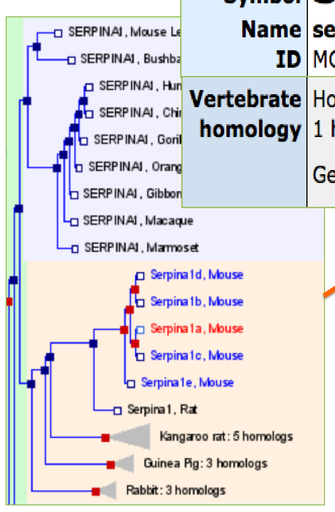
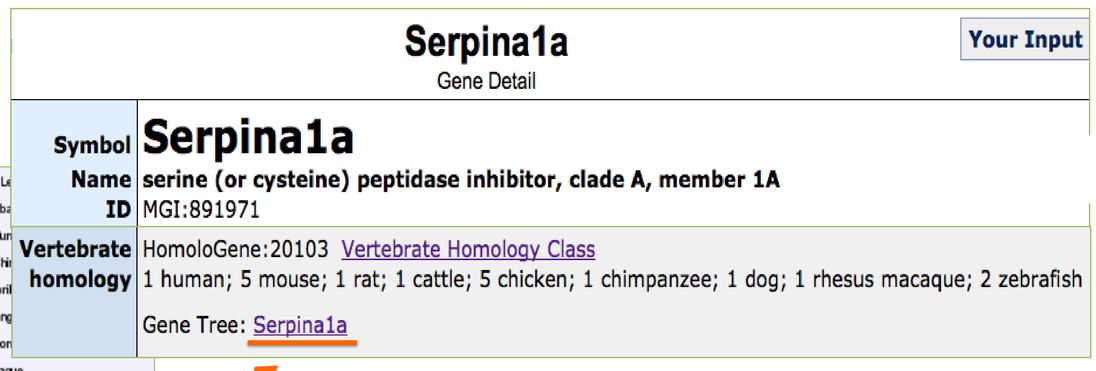
Assessment of comparative functional annotation propagation in mouse

Li Ni; Mary E. Dolan; Judith A. Blake
Bioinformatics and Computational Biology,
The Jackson Laboratory, 600 Main St., Bar Harbor, ME 04609
Mouse Genome Informatics - www.informatics.jax.org

In-Paralog Gene Duplications Lead to M:N Orthology Assertions

Branch of *Serpina1* GeneTree




Serpina1a
Gene Detail

Your Input

Symbol	**Serpina1a**
Name	**serine (or cysteine) peptidase inhibitor, clade A, member 1A**
ID	MGI:891971
Vertebrate homology	HomoloGene:20103 Vertebrate Homology Class 1 human; 5 mouse; 1 rat; 1 cattle; 5 chicken; 1 chimpanzee; 1 dog; 1 rhesus macaque; 2 zebrafish Gene Tree: Serpina1a

Gene Detail Page for *Serpina1a* with new Vertebrate Homology section shows one human gene is associated with five mouse genes. Vertebrate Homology now includes chicken and zebrafish orthologs.

Utilize M:N Orthology Assertions to make Inferential Gene Ontology (GO) Annotations

Symbol	**Gapdh**
Name	glyceraldehyde-3-phosphate dehydrogenase
ID	MGI:95640

Export: Text File | Excel File

Category	Classification Term	Evidence	Inferred From	Reference(s)
Molecular Function	glyceraldehyde-3-phosphate dehydrogenase (NAD+) (phosphorylating) activity	IMP	MGI:2	
Molecular Function	glyceraldehyde-3-phosphate dehydrogenase (NAD+) (phosphorylating) activity	ISO	RGD:2	
Molecular Function	microtubule binding	ISO	RGD:2661	J:155856
Molecular Function	peptidyl-cysteine S-nitrosylase activity	ISO	RGD:2661	J:155856
Cellular Component	cytoplasm	IDA		J:92551, J:142671, J:130677
Cellular Component	cytoplasm	ISO	RGD:2661	J:155856
Cellular Component	cytoplasm	ISO	P04406	J:164563
Cellular Component	cytosol	ISO	RGD:2661	J:155856
Cellular Component	extracellular vesicular exosome	ISO	P04406	J:164563
Cellular Component	lipid particle	ISO	P04406	J:164563
Cellular Component	microtubule cytoskeleton	ISO		
Cellular Component	mitochondrion	IDA		
Cellular Component	nucleus	ISO	R	
Cellular Component	plasma membrane	ISO	P04406	J:164563
Cellular Component	ribonucleoprotein complex	ISO	P04406	J:164563
Biological Process	cellular response to interferon-gamma	ISO	P04406	J:164563
Biological Process	multicellular organismal development	IMP	MGI:2183515	J:15477
Biological Process	negative regulation of translation	ISO	P04406	J:164563
Biological Process	oxidation-reduction process	IMP	MGI:2183515	J:15477

inferred mouse annotations from rat

inferred mouse annotations from human

M:N relationships are used to infer GO
Only do it for human/rat annotations that have Experimental evidence codes
- The inferred mouse annotations get ISO evidence code
- If human/rat-mouse are 1-1 do it for Function, Component, and Process
- If human/rat-mouse are not 1-1, do for F, C

For Mouse to Mouse (more than 1 mouse gene in set)
- Infer only from GO annotations that have experimental evidence codes
- Evidence code for the inferred annotations is ISO
- Do this only for Function and Cellular Component

M:N Inferred mouse annotations in MGI
8,023 genes: 44,907 annotations – Human
4,127 genes: 24,250 annotations – Rat

M:N Better Reflects Mouse Models and Their Relationships to Complex Human Diseases

Vertebrate Homology Class

Source	HomoloGene (Release 67, Dec 12, 2012)
Class ID	292

Comparative GO Graph (mouse, human, rat) HomoloGene:292 Multiple Sequence Alignment MGI Homology Information

Species	Symbol	Gene Links	Genetic Location	Genome Coordinates (mouse and human only)	Associated Human Diseases	Sequences (select all / deselect all / get FASTA / Go)
human	SMN1	HGNC:11117 (HGNC) 6606 (Entrez Gene) 600354 (OMIM)	Chr5 q13.2	Chr5:70220768-70248839 (+) GRCh37.p2	Spinal Muscular Atrophy, Type III; SMA3 Spinal Muscular Atrophy, Type II; SMA2 Spinal Muscular Atrophy, Type IV; SMA4	Q16637 (UniProt | EBI) NM_000344 (RefSeq)
	SMN2	HGNC:11118 (HGNC) 6607 (Entrez Gene) 601627 (OMIM)	Chr5 q13.2	Chr5:69345350-69373422 (+) GRCh37.p2	Spinal Muscular Atrophy, Type III; SMA3	Q16637 (UniProt | EBI) NM_017411 (RefSeq)
mouse	Smn1	MGI:109257 (MGI) 20595 (Entrez Gene) Gene Tree VISTA-Point	Chr13 52.99 cM	Chr13:100124852-100137690 (+) GRCm38	Spinal Muscular Atrophy, Type III; SMA3 Spinal Muscular Atrophy, Type II; SMA2 Spinal Muscular Atrophy, Type I; SMA1	OTTMUSG00000033496 (VEGA) OTTMUSP00000045374 (VEGA) OTTMUST00000084239 (VEGA)
rat	Smn1	64301 (Entrez Gene)	Chr2 q12			O35876 (UniProt | EBI)
cattle	SMN1	281492 (Entrez Gene)	Chr20 q14			
chicken	SMN	374025 (Entrez Gene)	ChrZ			
chimpanzee	SMN1	461829 (Entrez Gene)	Chr5			
dog	SMN	403896 (Entrez Gene)	Chr2			
rhesus macaque	SMN1	677703 (Entrez Gene)	Chr6			
zebrafish	smn1	30432 (Entrez Gene)	Chr5			

Human Disease and Mouse Model Detail

Human Disease	**Spinal Muscular Atrophy, Type III; SMA3** OMIM ID: 253400
Synonyms	Kugelberg-Welander Syndrome; KWS; Muscular Atrophy, Juvenile; SMA III; Spinal Muscular Atrophy, Mild Childhood and Adolescent Form
View all models	View ALL (2) mouse models for this human disease.
Genes and mouse models	Mutations in human and/or mouse homologs are associated with this disease *Gene is associated with the disease in this species

Mouse Homologs	Human Homologs	Mouse Models	Mouse : Human Homology Class
Smn1*	SMN1*, SMN2*	View 2 models	1:2 Homology

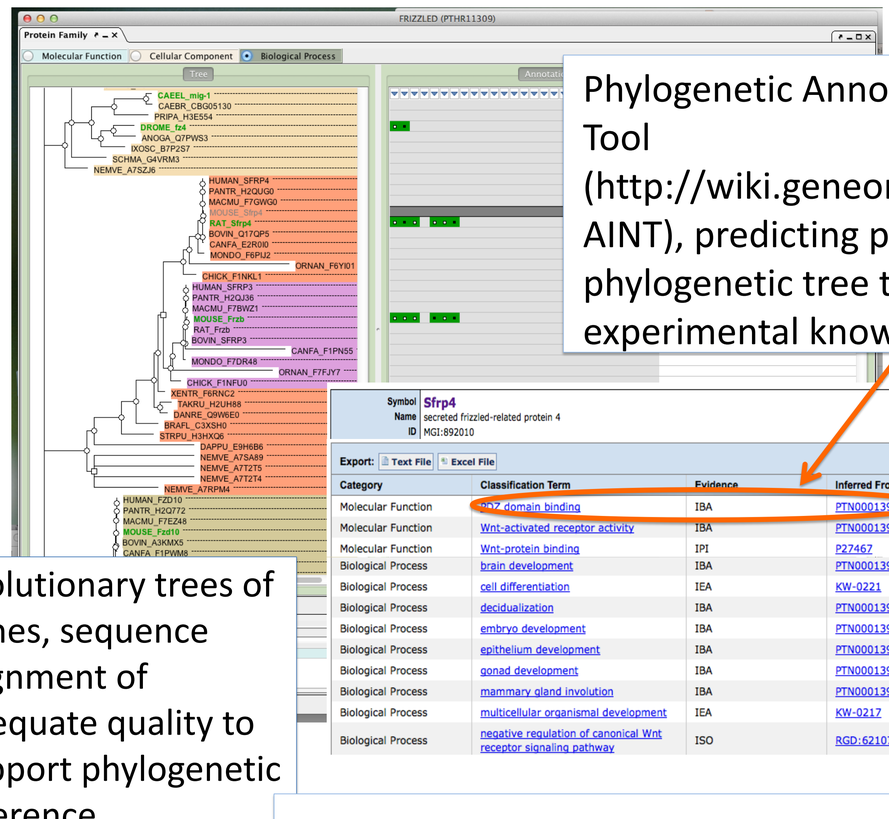
Mouse Models

Human Disease Modeled: Spinal Muscular Atrophy, Type III; SMA3
Associated Mouse Gene: Smn1

Allelic Composition	Genetic Background	Reference	Phenotypes
Smn1^{tm1Msd}/Smn1^{tm1Msd} Tg(SMN1*A2G)2023Ahmb/0 Tg(SMN2)89Ahmb/0	involves: 129P2/OlaHsd * FVB/N	J:81238	View
Smn1$^{tm1.1Dscd}$/Smn1$^{tm1.1Dscd}$	involves: 129S/SvEv * 129S4/SvJaeSor * C57BL/6	J:164889	View

Each associated human disease links to a Human Disease and Mouse Model Detail Page.

Other Inferential Annotations in MGI: Phylogenetic-Based Propagation



Phylogenetic Annotation and Inference Tool (http://wiki.geneontology.org/index.php AINT), predicting protein function using a phylogenetic tree together with available experimental knowledge of proteins.

Symbol	Sfrp4
Name	secreted frizzled-related protein 4
ID	MGI:892010

Export: Text File | Excel File

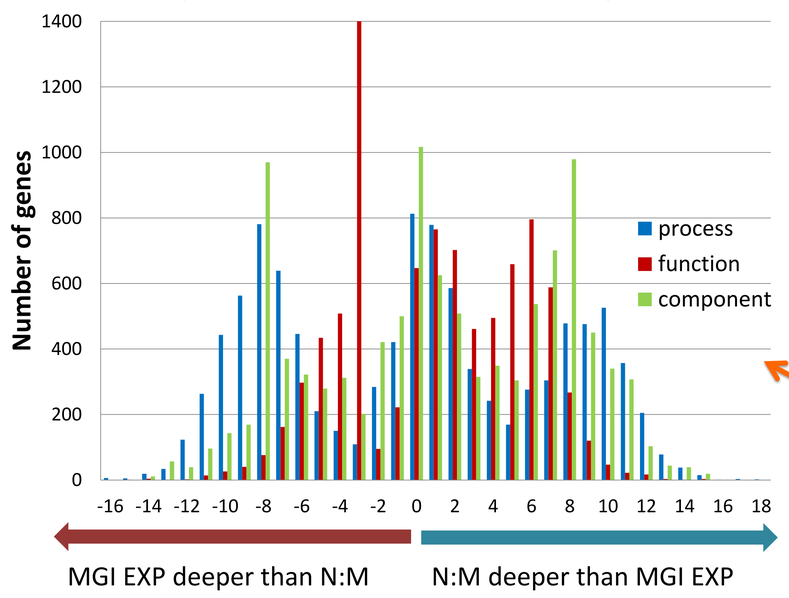
Category	Classification Term	Evidence	Inferred From	Reference(s)
Molecular Function	[illegible] domain binding	IBA	PTN000139394	J:161428
Molecular Function	Wnt-activated receptor activity	IBA	PTN000139394	J:161428
Molecular Function	Wnt-protein binding	IPI	P27467	J:121773
Biological Process	brain development	IBA	PTN000139394	J:161428
Biological Process	cell differentiation	IEA	KW-0221	J:60000
Biological Process	decidualization	IBA	PTN000139579	J:161428
Biological Process	embryo development	IBA	PTN000139394	J:161428
Biological Process	epithelium development	IBA	PTN000139573	J:161428
Biological Process	gonad development	IBA	PTN000139394	J:161428
Biological Process	mammary gland involution	IBA	PTN000139575	J:161428
Biological Process	multicellular organismal development	IEA	KW-0217	J:60000
Biological Process	negative regulation of canonical Wnt receptor signaling pathway	ISO	RGD:621075	J:155856

Evolutionary trees of genes, sequence alignment of adequate quality to support phylogenetic inference

PAINT Inferred mouse annotations in MGI
551 genes: total of 2720 annotations

Quality comparison of functional annotations

Quality comparison of MGI experimental annotations with N:M or PAINT inferred annotations using depth in the GO tree



The set includes all mouse genes that have either experimental or N:M inferred annotation.

The set includes only mouse genes that have been annotated using PAINT