### Gene Ontology (GO)

MCDB 187

# What is the Gene Ontology?

- Set of biological phrases (terms) which are applied to genes:
  - protein kinase
  - apoptosis
  - membrane

# What is the Gene Ontology?

- Genes are linked, or associated, with GO terms by trained curators at genome databases
  - known as 'gene associations' or GO annotations
- Some GO annotations created automatically

### What is the Gene Ontology?

 Allows biologists to make inferences across large numbers of genes without researching each one individually

#### How does GO work?

- GO is species independent
  - some terms, especially lower-level, detailed terms may be specific to a certain group
    - e.g. photosynthesis
  - But when collapsed up to the higher levels, terms are not dependent on species

#### How does GO work?

## What information might we want to capture about a gene product?

- What does the gene product do?
- Where and when does it act?
- Why does it perform these activities?

#### GO structure

• GO terms divided into three parts:

- cellular component
- molecular function
- biological process









• Enzyme complexes in the component ontology refer to places, not activities.

#### **Molecular Function**

activities or "jobs" of a gene product



glucose-6-phosphate isomerase activity



insulin binding

insulin receptor activity

#### **Molecular Function**

- A gene product may have several functions; a function term refers to a single reaction or activity, not a gene product.
- Sets of functions make up a biological process.

a commonly recognized series of events



#### cell division



transcription



#### regulation of gluconeogenesis



## Ontology Structure

- Terms are linked by two relationships
  - is-a 🚺
  - part-of (P)



membrane

chloroplast membrane

### Ontology Structure

- Ontologies are structured as a hierarchical directed acyclic graph (DAG)
- Terms can have more than one parent and zero, one or more children



#### mitochondrial membrane

chloroplast membrane

### Examples of GO structures



## Examples of GO structures

🖃 all : all ( 166775 ) 🗣 GO:0008150 : biological process (118690) GO:0009987 : cellular process (71171) GO:0050875 : cellular physiological process ( 65087 ) GO:0044237 : cellular metabolism ( 41108 ) GO:0006139: nucleobase, nucleoside, nucleotide and nucleic acid metabolism (16561) GO:0006259 : DNA metabolism (4671) GO:0006260 : DNA replication (1115) GO:0007582 : physiological process (73658) GO:0050875 : cellular physiological process ( 65087 ) GO:0044237 : cellular metabolism (41108) GO:0006139: nucleobase, nucleoside, nucleotide and nucleic acid metabolism (16561) GO:0006259 : DNA metabolism (4671) GO:0006260 : DNA replication (1115) GO:0008152 : metabolism (44953) GO:0044237 : cellular metabolism (41108) GO:0006139: nucleobase, nucleoside, nucleotide and nucleic acid metabolism (16561) GO:0006259 : DNA metabolism (4671) GO:0006260 : DNA replication (1115)
 GO:0043170 : macromolecule metabolism (23499) GO:0043283 : biopolymer metabolism (13529) GO:0006259 : DNA metabolism (4671) GO:0006260 : DNA replication (1115) GO:0044238 : primary metabolism (36601) GO:0006139: nucleobase, nucleoside, nucleotide and nucleic acid metabolism (16561) GO:0006259 : DNA metabolism (4671) GO:0006260 : DNA replication (1115)

#### GO tools

- GO resources are freely available to anyone to use without restriction
  - Includes the ontologies, gene associations and tools developed by GO
- Other groups have used GO to create tools for many purposes:

http://www.geneontology.org/GO.tools

#### AmiGO

000	AmiGO: BLAST Query						
( ) +	A http://amigo.geneontology.org/cgi-bin/amigo/blast.cgi?session_id=2224amigo1320209937 C Q- amigo	9	0				
⇔ [] =	IDP FIRE IPA Precipitation Blackboard DAVID Yahool Stock, OptE FINANCIAL Apple (150) * Digication Editorial Manager® UCLA/UCOP Reservation	Pellegrinilab GATHER Mila MCD8	20				
the Gene Ontology AmiGO							
	Search Browse BLAST Homolog Annotations Tools & Resources Help						
Search GO O terms O genes or proteins exact match Submit							

#### **BLAST Search**

The sequence search is performed using either BLASTP or BLASTX (from the WU-BLAST package), depending on the type of the input sequence.

BLAST Query					
Enter your query 😫					
Enter a UniProtKB accession or upload a text file of queries or paste in FASTA sequence(s)					
UniProtKB accession:					
Text file (maximum file size 500K): Choose File no file selected					
FASTA sequence(s):					
Sequences should be separated with an empty line.					
>pil61677879igbiAAX52480.1] actin SC, isoform D [Drosophila melanogaster] MCDIEVAAKVVONGSCMCKAGFAGDDAPRAVFPSIVCRPRHQCVNVGM GQKDSYVGDLAQSKRGLITLXY PBHCHTYNDDDDDCWPHETPHTYNELAVAPEDHPVLLTEAPLNIKANREKMT QMFETTNTPAMYVAIQAV LSLYASGKTTG/UQSCDCVSHTVPYYEGYALPHAILRI,DLAGRDLTDYLMKI LTRCYSTTTARRUY RDKICSLQYVALDFIQUMATAASSSLEXSYELPDGQYITICNEEFRCPEALF QPSFLGMLACGHETTY SSMKCDVDHKDLYANTVLSGGTTMYPCIADEMQKEITALAPSTMKIKIAP PERYSYMEGSLASL STFQQMWBKQEYDESGPSNHRKCF					
Maximum number of sequences: 100					

#### AmiGO

0.0	0 0 0 AmiGO: BLAST Results								
4	+ A http://a	migo.geneontology.org/cgi-bin/amigo/bla	ast.cgi		c Q.	amigo		•	
60	CO III IDP FIRE	IPA Precipitation Blackboard DAVID	Yahoo! Stock, OptiE FINANCIAL	Apple (150) * Digication	Editorial Manager®	UCLA/UCOP Reservation	Pellegrinilab CA	THER MILL MC	08
Q	uery Summa	ry							
You	ur job contains 1	sequence.							
Par	rameters								
Th	reshold: 0.1								
Ma	ximum number o	alignments shown: 50							
BL	AST filter: on								
0	uoni Coquer	~							_
Q	uery sequen	ce							
>gi	1616778791gb1AAX5	1480.11 octin 5C, isoform D [Drosoph	ila melanogaster]						
MCD	EEVAAL WONGSONCKA	FAGDDAPRAVEPSIVGRPRHQGVMVGMGQKDSYVG CMEXTINHTEYNELRVAREEHPVLLTEAPLNEXAN	DEAQ						
TQI	MFETFNTPAMYVAIQAVI	SLYASGRTTGIVLDSGDGVSHTVPIYEGYALPHAI	LRLD						
LAG	ROLTOYLMKILTERGYS	TTTAEREIVRDIKEKLCYVALDFEQEMATAASSSS	LEKS						
YEL	POGQVITIGNERFRCPE/	ALFQPSFLGMEACGIHETTYNSIMKCDVDIRKDLYA APSTMETETIAPPERCYSWEGGSTLASLSTEROM	NTVL WTSK						
QEY	DESGPSIVHRKCF								
									_
H	igh Scoring C	iene Products							
	Select all [Dear all Perform an action with this page's selected gene products								
		Symbol, full nam	ne		Infi	ormation		P value	
0	Act5C		BLAST r	natch # protein from	Drosophila melani	ogaster		2.7e-203	
	Actin SC		view associa	ations 🗢					
			BLAST with	ActSC +					
0	Act42A		BLAST n	match  protein from a	Drosophila melani	ogaster		7.1e-203	
	Actin 42A		BLAST with A	2900ns =					
	zgc:123289		BLAST r	natch + gene_produc	t from Danio reris	2		2.4e-202	
			view associa	itions +					
			BLAST with zgc:12	23289 🛥					
	P53478		BLAST n	match + protein from	Gallus gallus			9.4e-201	
	Actin, cytoplas	mic type 5	view associal BLAST with PS	100ns +					
	act.d		BLAST WITH PS	natch 1 name from Co.	enorhahditis class	104		1.20-200	
1.10	977-4		BLAST P	gene nom ca	enormationals elega	115		1.26-200	_

#### AmiGO

000	AmiGO: Act	SC Associations		16		
Image: Head of the second se	gp=F8:F8gn0000042&session Stock, OptIE FINANCIAL A	n_id=2224amigo132 pple (150) = Digicati	on Editorial Ma	5 (Q* amigo nager® UCLA/UCOP Reser	vation Pellegrinilab CAT	HER Mila MCDB
the Gene Ontology					Aı	miGO
Search Brown	ie BLAST Homolo	g Annotations	Tools & Reso	urces Help		
Search GO	• tern	ms 🔾 genes or p	oroteins 🗆 ex	act match Submit		
	Ac	:t5C				
Actin 5C						
protein from Drosophila melanogaster (fruit fly)						
Term associations 4	Gene product information	n 🕈 Peptide Sequ	ence + Seque	nce information +		
Term Associations						
Download all association information in: D gene association form	at D RDF-XML					
Filter associations displayed     Filter Associations     Ontology     Evidence Code     Diological process     cellular component     molecular function     BD     BD     BD						
Select all Dear all Perform an action with this page's selected terms	0 60					
Accession, Term		Ontology	Qualifier	Evidence	Reference	Assigned by
GO:0000910 : cytokinesis	1835 gene products	biological		IMP	PMID:15380073	FlyBase
	view in cree	process		IMP	PMID:15547975	FlyBase
	(83)	bisheded		IMP	PMID:16174742	FlyBase
GO:0007010 : cytoskeleton organization	view in tree	process		135	PMID:10908588	Hybase
G0:0002121 : inter-male aggressive behavior	66 gene products view in tree	biological process		IMP	PMID:19519879	FlyBase