Sequence Analysis: Lecture 20

- Ontologies
  1. Introduction to the Gene Ontology
  2. Gene Ontology annotations
  3. Editing the Gene Ontology
  4. Practical applications for the Gene Ontology
  5. The Gene Ontology as one of many biological ontologies

Most slides courtesy of Pascale Gaudet, dictyBase curator, Northwestern University, Chicago, IL
# GenBank

<table>
<thead>
<tr>
<th>Year</th>
<th>1982</th>
<th>2000</th>
<th>2015</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of basepairs</td>
<td>600,000</td>
<td>5,000,000,000</td>
<td>200,000,000,000</td>
</tr>
</tbody>
</table>

![Graph showing the growth of GenBank](image)
Genome Databases

- Mouse Genome Informatics
- FlyBase: *Drosophila*
- WormBase: *C. elegans*
- The *Arabidopsis* Information Resource
- dictyBase: *Dictyostelium discoideum*
- Saccharomyces Genome Database: Budding Yeast
- ZFIN: Zebrafish
- EcoGene - *E. coli*
  - GeneCards
  - Human ensembl
  - NCBI human genome resources

* manually curated by scientists
Gene Page for mhkA

dictyBase Home >> Chromosome 6:65301..70117 >> Gene: mhkA

General Information

Gene Name     mhkA
Synonym       MHCK-A
Gene Product   protein serine/threonine kinase
               myosin heavy chain kinase
               alpha kinase
Description    myosin II heavy chain kinase that contains an N-terminal coiled-coil domain, a central MHCK domain, and a C-terminal domain with 7 WD-40 repeats; drives myosin II disassembly
Name Description mhk = Myosin Heavy chain Kinase

Chromosomal Coordinates

Location      Chromosome 6 coordinates 65989 to 69429, Watson strand

Genome Browser Snapshot

Protein Information View mhkA Sequence Information Page

Length         1,146 aa
Molecular Weight 128,963 Da
Protein Domain  Predicted Protein Domains at GeneDB

Gene Ontology Annotations for mhkA View evidence and references

Molecular Function actin binding (IMP), myosin II binding (IPI), ATP binding (IPI), myosin-heavy-chain kinase activity (IDA), protein serine/threonine kinase activity (IDA), actin filament binding (IDA)
Biological Process autophosphorylation (IDA), protein amino acid phosphorylation (IDA), myosin II filament disassembly (IMP), myosin II filament disassembly (IDA)
Cellular Component cell cortex (IDA), cytoplasm (IDA)

Expression

Expression Profile

Phenotype View mhkA Phenotype details and references

Null
Increased myosin II assembly
Overexpression reduced growth in suspension, development arrests at mound stage, aberrant cytokinesis, reduced myosin II assembly
Too Much Published Literature

...to read in order to find out what your gene does....

- **PubMed**: db of over 15 million citations
  - Basic search:
    rad51 → 3308 articles
  - Organism Limited search:
    rad51 AND Human (organism) → 1737
  - Disease Limited search:
    rad51 AND cancer → 1535
Ontology

• Ontologies relate **facts** to **knowledge**

**facts**
- may be known/unknown/little known
- not attached to knowers
- unchanging

**knowledge**
- attached to knower
- may disappear
Gene Ontology

- Gene *annotation* system
- Controlled *vocabulary* that can be applied to all organisms
- Used to describe *gene products*
What is it?

• What is a cell?
Cell
Cell
Cell
What is it?

• The same **name** can be used to describe different **concepts**
What is it?
What is it?

• Glucose synthesis
• Glucose biosynthesis
• Glucose formation
• Glucose anabolism
• Gluconeogenesis

• All refer to the process of making glucose from simpler components
What is its name?

- The same *name* can be used to describe different *concepts*.
- The same *concept* can be described using different *names*.

→ Comparison is difficult – in particular across species or across databases.
What is the Gene Ontology?

A part of the solution:

- A controlled vocabulary that can be applied to all organisms

- Used to describe gene products - proteins and RNA - in any organism
Ontology

• In philosophy, the most fundamental branch of metaphysics. It studies **being** or **existence** as well as the basic categories thereof—trying to find out what **entities** and what **types of entities** exist.
  – Wikipedia

• Ontologies provide **controlled, consistent vocabularies** to describe concepts and relationships, thereby enabling **knowledge sharing**
  – Gruber 1993
Ontology

Includes:

1. Terms
2. Definitions
3. Logical relationships
4. Explicitness

Definition spectrum:
- Vague
- Explicit
Ontologies can be represented as graphs, where the **nodes** are connected by **edges**

- Nodes = **concepts** in the ontology
- Edges = **relationships** between the concepts
Ontology Structure

• The Gene Ontology is structured as a hierarchical directed acyclic graph (DAG)

• Terms can have more than one parent and zero, one or more children

• Terms are linked by two relationships
  – **is-a**
  – **part-of**
The value of any kind of data is greatly enhanced when it exists in a form that allows it to be integrated with other data. One approach to integration is through the annotation of multiple bodies of data using common controlled vocabularies or 'ontologies'. Unfortunately, the very success of this approach has led to a proliferation of ontologies, which itself creates obstacles to integration. The Open Biomedical Ontologies (OBO) consortium is pursuing a strategy to overcome this problem. Existing OBO ontologies, including the Gene Ontology, are undergoing coordinated reform, and new ontologies are being created on the basis of an evolving set of shared principles governing ontology development. The result is an expanding family of ontologies designed to be interoperable and logically well formed and to incorporate accurate representations of biological reality. We describe this OBO Foundry initiative and provide guidelines for those who might wish to become involved.
core relationships

- **OBO_REL:is_a**
  - Name *is_a*
  - Relation properties [transitive] [reflexive] [anti-symmetric]
  - Definition For continuants: C is_a C' if and only if: given any c that instantiates C at a time t, c instantiates C' at t. For processes: P is_a P' if and only if: that given any p that instantiates P, then p instantiates P'.

- **OBO_REL:part_of**
  - Name *part_of*
  - Relation properties [transitive] [reflexive] [anti-symmetric]
  - Definition For continuants: C part_of C' if and only if: given any c that instantiates C at a time t, there is some c' such that c' instantiates C' at time t, and c *part_of* c' at t. For processes: P part_of P' if and only if: given any p that instantiates P at a time t, there is some p' such that p' instantiates P' at time t, and p *part_of* p' at t. (Here *part_of* is the instance-level part-relation.)

--> .obo file format
Simple hierarchies (Trees)

- Single parent

Directed Acyclic Graphs

- One or more parents
Directed Acyclic Graphs

- Protein complex
- Organelle
- Mitochondrion
- Fatty acid beta-oxidation multienzyme complex

Other protein complexes
- Other organelles
True Path Rule

• The path from a child term all the way up to its top-level parent(s) must always be true

cell
  📢 cytoplasm
    📢 chromosome
      📢 nuclear chromosome
  📢 nucleus
    📢 nuclear chromosome
How does GO work?

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

- **What** does the gene product do?
- **Why** does it perform these activities?
- **Where** does it act?
GO: Three ontologies

What does it do?

What processes is it involved in?

Where does it act?

Molecular Function

Biological Process

Cellular Component

gene product
Cellular Component

- where a gene product acts
Mitochondrial membrane paths

- cellular_component
  - cell
    - intracellular
      - intracellular organelle
        - cytoplasm
      - organelle
        - membrane-bound organelle
        - mitochondrion
          - mitochondrial membrane
    - organelle
      - membrane-bound organelle
        - membrane
          - membrane-bound organelle

Number of genes: 7268, 5434, 4974, 3919, 3919, 3600, 942, 2556, 3600, 572, 382, 194
Biological Process
Molecular Function

- A single reaction or activity, not a gene product
- A gene product may have several functions
- Sets of functions make up a biological process
Molecular Function

[Diagram showing metabolic pathways including glucose 6-phosphate, phosphoenolpyruvate, pyruvate, lactate, oxaloacetate, and the Krebs cycle.]
Carbonate dehydratase activity
What’s in a GO term?

**term**: gluconeogenesis

**id**: GO:0006094

**definition**: The formation of glucose from noncarbohydrate precursors, such as pyruvate, amino acids and glycerol.
No GO Areas

• GO covers ‘normal’ functions and processes
  – No pathological processes
  – No experimental conditions
• NO evolutionary relationships
• NO gene products
• NOT a system of nomenclature
all : all (164672)

- GO:0008150: biological_process (116553)
  - GO:0007610: behavior (3333)
  - GO:0000004: biological process unknown (33114)
  - GO:0009987: cellular process (69630)
  - GO:0007275: development (11890)
  - GO:0040007: growth (3071)
  - GO:0044419: interaction between organisms (775)
  - GO:0007582: physiological process (74243)
  - GO:0043473: pigmentation (83)
  - GO:0050789: regulation of biological process (12939)
  - GO:0000003: reproduction (3821)
  - GO:0016032: viral life cycle (278)

- GO:0005575: cellular_component (102004)
  - GO:0005623: cell (72949)
    - GO:0008372: cellular component unknown (25487)
    - GO:0031012: extracellular matrix (566)
    - GO:0005576: extracellular region (4320)
    - GO:0043226: organelle (53431)
    - GO:0043234: protein complex (10154)
    - GO:0019012: virion (134)

- GO:0003674: molecular_function (114384)
  - GO:0016209: antioxidant activity (457)
  - GO:0005488: binding (29930)
  - GO:0003824: catalytic activity (38141)
  - GO:0030188: chaperone regulator activity (40)
  - GO:0030234: enzyme regulator activity (1806)
    - GO:0005554: molecular function unknown (34043)
  - GO:0003774: motor activity (487)
    - GO:0045735: nutrient reservoir activity (45)
    - GO:0031386: protein tag (12)
  - GO:0004871: signal transducer activity (8305)
  - GO:0005198: structural molecule activity (3366)
  - GO:0030528: transcription regulator activity (7662)
  - GO:0045182: translation regulator activity (687)
  - GO:0005215: transporter activity (10021)
    - GO:0030533: triplet codon-amino acid adaptor activity (1217)
## Content of GO

<table>
<thead>
<tr>
<th>Category</th>
<th>Terms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Molecular Function</td>
<td>7,309 terms</td>
</tr>
<tr>
<td>Biological Process</td>
<td>10,041 terms</td>
</tr>
<tr>
<td>Cellular Component</td>
<td>1,629 terms</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>18,975 terms</strong></td>
</tr>
</tbody>
</table>

**Definitions:** 94.9%

**Obsolete terms:** 992

*As of October 2005*
Outline

1. Introduction to the Gene Ontology
2. Gene Ontology annotations
3. Editing the Gene Ontology
4. Practical applications for the Gene Ontology
5. The Gene Ontology as one of many biological ontologies
Annotation of gene products with GO terms

Mitochondrial P450
Cellular component: mitochondrial inner membrane
GO:0005743

Biological process: Electron transport
GO:0006118

Molecular function: monooxygenase activity
GO:0004497

substrate + O_2 = CO_2 + H_2O product
Other gene products annotated to **monooxygenase activity** (GO:0004497)

- monooxygenase, DBH-like 1   (mouse)
- prostaglandin I2 (prostacyclin) synthase (mouse)
- flavin-containing monooxygenase (yeast)
- ferulate-5-hydrolase 1   (arabidopsis)
Two types of GO Annotations:

- Electronic Annotation
- Manual Annotation

All annotations **must**:

- be attributed to a source
- indicate what evidence was found to support the GO term-gene/protein association
Manual Annotations

• High–quality, specific gene/gene product associations made, using:
  • Peer-reviewed papers
  • Evidence codes to grade evidence

BUT – is very time consuming and requires trained biologists
Electronic Annotations

- Provides large-coverage
- High-quality

**BUT** – annotations tend to use high-level GO terms and provide little detail.
Electronic Annotations: Methods

1. Database entries
   - Manual mapping of GO terms to concepts external to GO (‘translation tables’)
   - Proteins then electronically annotated with the relevant GO term(s)

2. Automatic sequence similarity analyses to transfer annotations between highly similar gene products
Electronic Annotations

Fatty acid biosynthesis (Swiss-Prot Keyword) → GO:Fatty acid biosynthesis (GO:0006633)

EC:6.4.1.2 (EC number) → GO:acetyl-CoA carboxylase activity (GO:0003989)

IPR000438: Acetyl-CoA carboxylase carboxyl transferase beta subunit (InterPro entry) → GO:acetyl-CoA carboxylase activity (GO:0003989)
Mappings of external concepts to GO

<table>
<thead>
<tr>
<th>Database</th>
<th>Index File</th>
</tr>
</thead>
<tbody>
<tr>
<td>Swiss-Prot</td>
<td>spkw2go</td>
</tr>
<tr>
<td>Enzyme Commission</td>
<td>ec2go</td>
</tr>
<tr>
<td>EGAD</td>
<td>egad2go</td>
</tr>
<tr>
<td>GenProtEC</td>
<td>genprotec2go</td>
</tr>
<tr>
<td>TIGR role</td>
<td>tigr2go</td>
</tr>
<tr>
<td>TIGR Families</td>
<td>tigrfams2go</td>
</tr>
<tr>
<td>InterPro</td>
<td>interpro2go</td>
</tr>
<tr>
<td>MIPS Funcat</td>
<td>mips2go</td>
</tr>
<tr>
<td>MetaCyc Pathways</td>
<td>metacyc2go</td>
</tr>
<tr>
<td>MultiFun Classifications</td>
<td>multifun2go</td>
</tr>
<tr>
<td>Pfam Domains</td>
<td>pfam2go</td>
</tr>
<tr>
<td>Prodom Domains</td>
<td>prodom2go</td>
</tr>
<tr>
<td>Prints Domains</td>
<td>prints2go</td>
</tr>
<tr>
<td>ProSite Domains</td>
<td>prosite2go</td>
</tr>
<tr>
<td>Smart Domains</td>
<td>smart2go</td>
</tr>
</tbody>
</table>

EC:1.1.1.1  >  GO:alcohol dehydrogenase activity ; GO:0004022
EC:1.1.1.10  >  GO:L-xylulose reductase activity ; GO:0050038
EC:1.1.1.104  >  GO:4-oxoproline reductase activity ; GO:0016617
EC:1.1.1.105  >  GO:retinol dehydrogenase activity ; GO:0004745
Manual Annotations: Methods

1. Extract information from published literature

2. Curators performs manual sequence similarity analyses to transfer annotations between highly similar gene products (BLAST, protein domain analysis)
In this study, we report the isolation and molecular characterization of the *B. napus* PERK1 cDNA, that is predicted to encode a novel receptor-like kinase. We have shown that like other plant RLKs, the kinase domain of PERK1 has serine/threonine kinase activity, In addition, the location of a PERK1-GFP fusion protein to the plasma membrane supports the prediction that PERK1 is an integral membrane protein. These kinases have been implicated in early stages of wound response...

PubMed ID: 12374299

**Function:** protein serine/threonine kinase activity  
**GO:**0004674

**Component:** integral to plasma membrane  
**GO:**0005887

**Process:** response to wounding  
**GO:**0009611

Thought: Can we do this by natural language processing?
Additional points

• A gene product can have several functions, cellular locations and be involved in many processes

• Annotation of a gene product to one ontology is independent from its annotation to other ontologies

• Annotations are only to terms reflecting a normal activity or location

• Usage of ‘unknown’ GO terms
Unknown vs. Unannotated

• “Unknown” is used when the curator has determined that there is no existing literature to support an annotation.
  – Biological process unknown GO:0000004
  – Molecular function unknown GO:0005554
  – Cellular component unknown GO:0008372

• Annotation “unknown” is NOT the same as having no annotation.
  – No annotation means that no one has looked yet. *Unknown* means they looked and did not find.
## GO Evidence Codes

<table>
<thead>
<tr>
<th>Code</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>IEA</td>
<td>Inferred from <strong>Electronic Annotation</strong></td>
</tr>
<tr>
<td>NAS</td>
<td><strong>Non-traceable Author Statement</strong></td>
</tr>
<tr>
<td>TAS</td>
<td><strong>Traceable Author Statement</strong></td>
</tr>
<tr>
<td>ND</td>
<td><strong>No Data</strong> Use with annotation to unknown</td>
</tr>
<tr>
<td>IDA</td>
<td>Inferred from <strong>Direct Assay</strong> Manually annotated</td>
</tr>
<tr>
<td>*IPI</td>
<td>Inferred from <strong>Physical Interaction</strong></td>
</tr>
<tr>
<td>*IGI</td>
<td>Inferred from <strong>Genetic Interaction</strong></td>
</tr>
<tr>
<td>IMP</td>
<td>Inferred from <strong>Mutant Phenotype</strong></td>
</tr>
<tr>
<td>IEP</td>
<td>Inferred from <strong>Expression Pattern</strong></td>
</tr>
<tr>
<td>*IC</td>
<td>Inferred from <strong>Curator</strong></td>
</tr>
<tr>
<td>*ISS</td>
<td>Inferred from <strong>Sequence Similarity</strong></td>
</tr>
</tbody>
</table>
## GO Evidence Codes

<table>
<thead>
<tr>
<th>Code</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>*IEA</td>
<td>Inferred from Electronic Annotation</td>
</tr>
<tr>
<td>IDA</td>
<td>Inferred from Direct Assay</td>
</tr>
<tr>
<td>IEP</td>
<td>Inferred from Expression Pattern</td>
</tr>
<tr>
<td>*IGI</td>
<td>Inferred from Genetic Interaction</td>
</tr>
<tr>
<td>IMP</td>
<td>Inferred from Mutant Phenotype</td>
</tr>
<tr>
<td>*IPI</td>
<td>Inferred from Physical Interaction</td>
</tr>
<tr>
<td>*ISS</td>
<td>Inferred from Sequence Similarity</td>
</tr>
<tr>
<td>TAS</td>
<td>Traceable Author Statement</td>
</tr>
<tr>
<td>NAS</td>
<td>Non-traceable Author Statement</td>
</tr>
<tr>
<td>*IC</td>
<td>Inferred from Curator</td>
</tr>
<tr>
<td>RCA</td>
<td>Inferred from Reviewed Computational Analysis</td>
</tr>
<tr>
<td>ND</td>
<td>No Data</td>
</tr>
</tbody>
</table>

**Manually annotated**

**IDA:**
- Enzyme assays
- *In vitro* reconstitution (transcription)
- Immunofluorescence
- Cell fractionation

**TAS:**
- In the literature source the original experiments referred to are traceable (referenced).
GO Evidence Codes: with/from

Additional information required for certain evidence codes

<table>
<thead>
<tr>
<th>Code</th>
<th>Definition</th>
<th>IGI:</th>
</tr>
</thead>
<tbody>
<tr>
<td>*IEA</td>
<td>Inferred from <strong>Electronic Annotation</strong></td>
<td>• a gene identifier for the &quot;other&quot; gene involved in the interaction</td>
</tr>
<tr>
<td>IDA</td>
<td>Inferred from <strong>Direct Assay</strong></td>
<td>*With column required</td>
</tr>
<tr>
<td>IEP</td>
<td>Inferred from <strong>Expression Pattern</strong></td>
<td>*IGI:</td>
</tr>
<tr>
<td>*IGI</td>
<td>Inferred from <strong>Genetic Interaction</strong></td>
<td>• a gene or protein identifier for the &quot;other&quot; protein involved in the interaction</td>
</tr>
<tr>
<td>IMP</td>
<td>Inferred from <strong>Mutant Phenotype</strong></td>
<td>IPI:</td>
</tr>
<tr>
<td>*IPI</td>
<td>Inferred from <strong>Physical Interaction</strong></td>
<td>• a gene or protein identifier for the &quot;other&quot; protein involved in the interaction</td>
</tr>
<tr>
<td>*ISS</td>
<td>Inferred from <strong>Sequence Similarity</strong></td>
<td>IC:</td>
</tr>
<tr>
<td>TAS</td>
<td>Traceable <strong>Author Statement</strong></td>
<td>• GO term from another annotation used as the basis of a curator inference</td>
</tr>
<tr>
<td>NAS</td>
<td>Non-traceable <strong>Author Statement</strong></td>
<td>ND:</td>
</tr>
<tr>
<td>*IC</td>
<td>Inferred from <strong>Curator</strong></td>
<td>No Data</td>
</tr>
<tr>
<td>RCA</td>
<td>Inferred from <strong>Reviewed Computational Analysis</strong></td>
<td></td>
</tr>
<tr>
<td>ND</td>
<td><strong>No Data</strong></td>
<td></td>
</tr>
</tbody>
</table>
Evidence code, confidence

TAS/IDA
IMP/IGI/IPI
ISS/IEP
NAS
IEA
Modifying the interpretation of an annotation: the Qualifier column

1. **NOT**
   - a gene product is NOT associated with the GO term
   - to document conflicting claims in the literature.

2. **Contributes to**
   - distinguishes between individual subunit functions and whole complex functions
   - used with GO Function Ontology

3. **Colocalizes with**
   - transiently or peripherally associated with an organelle or complex
   - used with GO Component Ontology
Annotation of a genome

- GO annotations are always work in progress
- Part of normal curation process
  - More specific information
  - Better evidence code
- Replace obsolete terms
- “Last reviewed” date
How to access the Gene ontology and its annotations

1. Downloads
   - Ontologies
   - Annotations: Gene association files
   - Ontologies and Annotations

2. Web-based access
   - AmiGO
     (http://www.godatabase.org)
   - QuickGO
     (http://www.ebi.ac.uk/ego)

among others...


## Gene Association Files


Ontology and annotation data is integrated in the mySQL and XML files available from the [GO Database site](http://www.geneontology.org).

**Notes:**
1. The files are compressed using the UNIX gzip utility; use the "Download" link to download the compressed file to your disk.
2. Where available (e.g. for the Compugen and GO Annotations at EBI files), please also see the appropriate README file.

<table>
<thead>
<tr>
<th>Biological Process</th>
<th>Molecular Function</th>
<th>Cellular Component</th>
<th>Total Gene Products Associated</th>
<th>Total References Included as Evidence</th>
<th>TAB Delimited File of Associations &amp; Last Update</th>
</tr>
</thead>
<tbody>
<tr>
<td>SGD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Saccharomyces cerevisiae</td>
<td>6454</td>
<td>6454</td>
<td>6437</td>
<td>6437</td>
<td>6437</td>
</tr>
<tr>
<td>FlyBase</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Drosophila melanogaster</td>
<td>9143</td>
<td>5835</td>
<td>9277</td>
<td>7696</td>
<td>6447</td>
</tr>
<tr>
<td>MGI</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mus musculus</td>
<td>12773</td>
<td>8377</td>
<td>13701</td>
<td>8781</td>
<td>12998</td>
</tr>
<tr>
<td>TAIR</td>
<td>11647</td>
<td>11644</td>
<td>6232</td>
<td>6232</td>
<td>20834</td>
</tr>
</tbody>
</table>
Anatomy of a gene association file

<table>
<thead>
<tr>
<th>Column</th>
<th>Content</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>DB</td>
<td>SGD, MGI</td>
</tr>
<tr>
<td>2</td>
<td>DB_Object ID</td>
<td>MGI:1234568</td>
</tr>
<tr>
<td>3</td>
<td>DB_Object_Symbol</td>
<td>Gras3</td>
</tr>
<tr>
<td>4</td>
<td>GO_ID Qualifier</td>
<td>NOT, co_localizes_with, contributes_to</td>
</tr>
<tr>
<td>5</td>
<td>GO_ID</td>
<td>GO:0001515</td>
</tr>
<tr>
<td>6</td>
<td>DB_Ref</td>
<td>PMID:234567</td>
</tr>
<tr>
<td>7</td>
<td>Evidence_Code</td>
<td>IDA, etc.</td>
</tr>
<tr>
<td>8</td>
<td>With/From</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>GO_aspect</td>
<td>P (process), C (component) F (function)</td>
</tr>
<tr>
<td>10</td>
<td>DB_Object_Name</td>
<td>Grasshopper 3 homlog</td>
</tr>
<tr>
<td>11</td>
<td>DB_Object_Synonym</td>
<td>Locust III, 0122345E12Rik</td>
</tr>
<tr>
<td>12</td>
<td>DB_Object_Type</td>
<td>Gene, transcript, or protein</td>
</tr>
<tr>
<td>13</td>
<td>Taxon</td>
<td>taxon:4932</td>
</tr>
<tr>
<td>14</td>
<td>Date</td>
<td>20050101</td>
</tr>
<tr>
<td>15</td>
<td>Assigned_by</td>
<td>DB (usually same as column 1)</td>
</tr>
</tbody>
</table>
Viewing Annotations

- **Amigo Browser:**
  [http://amigo.geneontology.org/amigo](http://amigo.geneontology.org/amigo)
  
  - A GO browser that tracks contributed GO annotations across species.
  
  - Uses annotation sets supplied in a specific format.
GO:0008150: biological_process (131343)
  ▪ GO:0007610: behavior (4249)
  ▪ GO:0000004: biological_process unknown (33366)
  ▪ GO:0009987: cellular process (82401)
  ▪ GO:0007275: development (18484)
▪ GO:0007582: physiological_process (87218)
    ▪ GO:0008218: bioluminescence (9)
    ▪ GO:0050875: cellular physiological process (73662)
    ▪ GO:0050817: coagulation (169)
    ▪ GO:0016265: death (2782)
    ▪ GO:0043062: extracellular structure organization and biogenesis (432)
    ▪ GO:0042592: homeostasis (1229)
    ▪ GO:0044419: interaction between organisms (856)
    ▪ GO:0051179: localization (19242)
    ▪ GO:0008152: metabolism (50429)
    ▪ GO:0050874: organismal physiological process (8717)
    ▪ GO:0015979: photosynthesis (105)
  ▪ GO:0050791: regulation of physiological process (14115)
  ▪ GO:0050896: response to stimulus (14316)
  ▪ GO:0048511: rhythmic process (377)
▪ GO:0046903: secretion (8402)
  ▪ GO:0046717: acid secretion (9)
  ▪ GO:0045457: ecdysteroid secretion (1)
    ▪ GO:007589: fluid secretion (26)
  ▪ GO:0046879: hormone secretion (88)
  ▪ GO:0007269: neurotransmitter secretion (414)
    ▪ GO:0030157: pancreatic juice secretion (3)
  ▪ GO:0009306: protein secretion (422)
  ▪ GO:0051046: regulation of secretion (108)
  ▪ GO:0046541: saliva secretion (5)
  ▪ GO:0045045: secretory pathway (7868)
  ▪ GO:0010162: seed dormancy (6)
  ▪ GO:0050789: regulation of biological process (17673)
  ▪ GO:0016032: viral life cycle (300)
▪ GO:0005575: cellular_component (121894)
▪ GO:0003674: molecular_function (129566)
secretion

Accession: GO:0046903
Ontology: biological_process
Synonyms: narrow: SAR small monomeric GTPase activity
Definition:
The regulated release of a substance by a cell or group of cells.
Comment: None

Term Lineage

all : all (5130)
  ⊞ GO:0006150 : biological_process (3614)
  ⊞ GO:0007582 : physiological_process (3088)
  ⊞ GO:0051179 : localization (457)
  ⊞ GO:0051234 : establishment_of_localization (452)
  ⊞ GO:0046903 : secretion (91)

External References

- COG (1)
- InterPro (1)
- Pfam (1)

All Gene Product Associations

(98 results)

Filter Associations

Datasource | Evidence Code | Species
---|---|---
All | All Curator Approved | All
FlyBase | IC | A. aeolicus
SGD | IMP | A. fulgidus

Qualifier | Symbol | Information | Source | Evidence | Reference
---|---|---|---|---|---
Anxa6 | annexin A6, gene from Rattus norvegicus | RGD | TAS | RGD:724802
Unc13h4 | Munc13-4 protein, gene from Rattus norvegicus | RGD | IEP | RGD:634401
Atp4a | ATPase, H+/K+ transporting, alpha polypeptide, gene from Rattus norvegicus | RGD | TAS | RGD:631722
Atp4b | ATPase, H+/K+ transporting, beta polypeptide, gene from Rattus norvegicus | RGD | TAS | RGD:631715
Cart | cocaine and amphetamine regulated transcript, gene from Rattus norvegicus | RGD | TAS | RGD:1309214
Anxa6

Information:
- annexin A6
- gene from *Rattus norvegicus*

Source:
- RGD

Reference:
- RGD:724802

Abstract:

CRHBP-30 is a member of the tumor protein DSG-protein family that was recently shown to regulate Ca2+-sensitive cation-transport in pancreatic acinar cells. Thomas, D.D., Hsiao, H.H., Taff, J.B., Wang, W., Fedorkow, L.A., and Gubins, G.E. (2001). J. Biol. Chem. 276, 20611-20617. In the present study, the Ca2+-sensitive phospholipid-binding protein annexin VI was purified from rat pancreas as a CRHBP-30-binding protein. The interaction between CRHBP-30 and annexin VI was demonstrated by cross-linking experiments and gel-filtration analysis and was shown to require the calcium-binding domain of annexin VI. In a calcium-dependent, dose-dependent manner, annexin VI inhibited CRHBP-30-mediated exocytosis of amylase from isolated acinar cells for 0-6 h, with the inhibitory effect being enhanced by the co-localization of CRHBP-30 and annexin VI in regions of acute immobility below the acinar plasma membrane. These results suggest that CRHBP-30 is a pancreatic membrane protein that is highly enriched in smooth muscle fractions of pancreas. Further, the content of CRHBP-30 in microsomes was significantly reduced in parallel acini obtained from rats that had been infused with a secretory dose of cholecystokinin for 4 h, demonstrating that secretagogue stimulation transiently alters the association of CRHBP-30 with membranes in cells. Collectively, the Ca2+-dependent binding of CRHBP-30 and annexin VI, together with their co-localization in the acinar cytoplasm, is consistent with a role for these molecules in acinar cell membrane trafficking events that are essential for digestive enzyme secretion.

Show data curated from this reference
**secretion**

**Accession:** GO:0046903  
**Aspect:** biological_process  
**Synonyms:** None  
**Definition:**  
The regulated release of a substance by a cell or group of cells.

**Term Lineage**

```
all : all (<201746 )
  GO:00068150 : biological_process (<131343 )
  GO:0007582 : physiological_process (<87218 )
  GO:0046903 : secretion (<8402 )
```

**External References**

- COG (1)  
- InterPro (1)  
- Pfam (1)

**Direct Gene Product Associations**

**Filter Associations**

<table>
<thead>
<tr>
<th>Datasource</th>
<th>Evidence Code</th>
<th>Species</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>All Curator Approved</td>
<td>H. pylori J99</td>
</tr>
<tr>
<td>FlyBase</td>
<td>IMP</td>
<td>H. sapiens</td>
</tr>
<tr>
<td>SGD</td>
<td>IGI</td>
<td>H. sp. NRC-1</td>
</tr>
</tbody>
</table>

**Gene Symbol**

- **KPCI_HUMAN**  
  **Type:** protein  
  **Datasource:** UniProt  
  **Evidence:** ISS  
  **Full Name:** Protein kinase C, iota type

- **Q8WVW0_HUMAN**  
  **Type:** protein  
  **Datasource:** UniProt  
  **Evidence:** NAS  
  **Full Name:** Protein kinase C, iota

- **RMS1_HUMANATGCC / G0st**  
  **Type:** protein  
  **Datasource:** UniProt  
  **Evidence:** NAS  
  **Full Name:** Regulating synaptic membrane exocytosis protein 1
<table>
<thead>
<tr>
<th>Gene Symbol</th>
<th>Type</th>
<th>Datasource</th>
<th>Evidence</th>
<th>Full Name</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>secretion</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KPCI_HUMAN ATG6CC/G0st</td>
<td>protein</td>
<td>UniProt</td>
<td>ISS</td>
<td>Protein kinase C, iota type</td>
</tr>
<tr>
<td>Q6WV0G_HUMAN</td>
<td>protein</td>
<td>UniProt</td>
<td>NAS</td>
<td>Protein kinase C, iota</td>
</tr>
<tr>
<td>RIMS1_HUMAN ATG6CC/G0st</td>
<td>protein</td>
<td>UniProt</td>
<td>NAS</td>
<td>Regulating synaptic membrane exocytosis protein 1</td>
</tr>
<tr>
<td><strong>acid secretion</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q6UG9G_HUMAN ATG6CC/G0st</td>
<td>protein</td>
<td>UniProt</td>
<td>IDA</td>
<td>Carnitine transporter 2</td>
</tr>
<tr>
<td><strong>arachidonic acid secretion</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q6GZQ4_HUMAN</td>
<td>protein</td>
<td>UniProt</td>
<td>IDA</td>
<td>Neuromedin U receptor 2</td>
</tr>
<tr>
<td><strong>fluid secretion</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AG2R_HUMAN ATG6CC/G0st</td>
<td>protein</td>
<td>UniProt</td>
<td>NR</td>
<td>Type-I angiotensin II receptor</td>
</tr>
<tr>
<td>ANFB_HUMAN ATG6CC/G0st</td>
<td>protein</td>
<td>UniProt</td>
<td>TAS</td>
<td>Natriuretic peptides B precursor [Contains: Gamma-brain natriuretic peptide; Brain natriuretic peptide 32]</td>
</tr>
<tr>
<td>ANPRA_HUMAN ATG6CC/G0st</td>
<td>protein</td>
<td>UniProt</td>
<td>TAS</td>
<td>Atrial natriuretic peptide receptor A precursor</td>
</tr>
<tr>
<td>GUC1B_HUMAN ATG6CC/G0st</td>
<td>protein</td>
<td>UniProt</td>
<td>TAS</td>
<td>Guanylyl cyclase activating protein 2</td>
</tr>
<tr>
<td>NPT2_HUMAN ATG6CC/G0st</td>
<td>protein</td>
<td>UniProt</td>
<td>TAS</td>
<td>Renal sodium-dependent phosphate transport protein 2</td>
</tr>
</tbody>
</table>
Querying the GO

AmiGO

Search GO

- all : all (201746)
- GO:0008150 : biological_process (131343)
- GO:0005575 : cellular_component (121894)
- GO:0003674 : molecular_function (129586)
- obsolete_biological_process : obsolete_biological_process (71)
- obsolete cellular_component (36)
- obsolete_molecular_function (941)

Search for GO terms or by Gene symbol/name

Filter queries by organism, data source or evidence

Last updated: 2005-05-02
<table>
<thead>
<tr>
<th>Go Term</th>
<th>Aspect</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>mitochondrion</td>
<td>cellular_component</td>
<td>A semi-autonomous, self-replicating organelle that occurs in varying number of cells.</td>
</tr>
<tr>
<td>mitochondrion organization and biogenesis</td>
<td>biological_process</td>
<td>The assembly and arrangement of a mitochondrion, including mitochondrial biogenesis.</td>
</tr>
<tr>
<td>axon transport of mitochondrion</td>
<td>biological_process</td>
<td>The directed movement of mitochondria along microtubules in nerve cell axons.</td>
</tr>
<tr>
<td>sperm mitochondrion organization and biogenesis</td>
<td>biological_process</td>
<td>The assembly and arrangement of sperm mitochondria, the process by which sperm mature.</td>
</tr>
<tr>
<td>mitochondrion targeting sequence binding</td>
<td>molecular_function</td>
<td>Interacting selectively with a mitochondrion targeting sequence, a specific protein.</td>
</tr>
<tr>
<td>mitochondrion transport along microtubule</td>
<td>biological_process</td>
<td>The directed movement of a mitochondrion along a microtubule, mediated by motor proteins.</td>
</tr>
<tr>
<td>mitochondrion distribution</td>
<td>biological_process</td>
<td>Spatial arrangement of mitochondria between and within cells.</td>
</tr>
<tr>
<td>mitochondrion positioning within cell</td>
<td>biological_process</td>
<td>Spatial arrangement of mitochondria within the cell.</td>
</tr>
<tr>
<td>protein insertion into mitochondrial membrane</td>
<td>biological_process</td>
<td>The process by which a protein is incorporated into a mitochondrial membrane.</td>
</tr>
<tr>
<td>protein insertion into mitochondrial membrane</td>
<td>biological_process</td>
<td>The process by which a protein is incorporated into a mitochondrial membrane.</td>
</tr>
<tr>
<td>Gene Product</td>
<td>Type</td>
<td>Datasource</td>
</tr>
<tr>
<td>--------------</td>
<td>------</td>
<td>------------</td>
</tr>
<tr>
<td>ERCA1_HUMAN</td>
<td>protein</td>
<td>UniProt</td>
</tr>
<tr>
<td>ERCA2_HUMAN</td>
<td>protein</td>
<td>UniProt</td>
</tr>
</tbody>
</table>
Entrez Gene

AMID: apoptosis-inducing factor (AIF)-like mitochondrion-associated inducer of death

Gene ID: 8483
Local tag: HGNC21411

Official Symbol: AMID and Name: apoptosis-inducing factor (AIF)-like mitochondrion-associated inducer of death provided by HGNC

Gene Nomenclature Committee

Transcripts and products: (shown on reverse complement genome) RefSeq below

Genomic context: Chromosome 10, Maps: 10q22.1

Gene type: protein coding
Gene name: AMID
Gene description: apoptosis-inducing factor (AIF)-like
RefSeq status: Reviewed

Organism: Homo sapiens
Lineage: Eukaryota; Metazoa; Chordata; Craniata; Cetartiodactyla; Primates; Homo

Gene aliases: PRG5

Summary: The protein encoded by this gene has significant homology to the apoptosis-inducing factor (AIF) protein. Overexpression of this gene has been shown to suppress protein p53 in colon cancer cells.

Bibliography:

Gene Ontology

Function: quinolide oxidoreductase activity, electron-transport flavoprotein dehydrogenase activity

Process: apoptotic mitochondrial changes, chromosome condensation, electron transport, induction of apoptosis, metabolism

Component: cytosol, mitochondrial outer membrane

Homology: Mouse, Rat

### General Information about the UniProt/Swiss-Prot entry

<table>
<thead>
<tr>
<th>Entry name</th>
<th>AATM_RABIT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary accession number</td>
<td>P12345</td>
</tr>
<tr>
<td>Entered in Swiss-Prot</td>
<td>Release 12, 01-OCT-1989</td>
</tr>
<tr>
<td>Sequence was last modified</td>
<td>Release 12, 01-OCT-1989</td>
</tr>
<tr>
<td>Annotations were last modified</td>
<td>Release 44, 05-JUL-2004</td>
</tr>
</tbody>
</table>

### Protein description

<table>
<thead>
<tr>
<th>Protein name</th>
<th>Aspartate aminotransferase</th>
</tr>
</thead>
<tbody>
<tr>
<td>EC</td>
<td>EC 2.6.1.1</td>
</tr>
<tr>
<td>Synonyms</td>
<td>Transaminase A, Glutamate oxaloacetate transaminase</td>
</tr>
</tbody>
</table>

### Origin of the protein

| Gene                           | Gene name | GOT2 |
|--------------------------------|------------|
| From                           | Oryctolagus cuniculus (Rabbit) |
| Taxonomy                       | Eukaryota; Metazoa; Chordata; Lagonomorpha; Lepidoeida; Oryctolagus cuniculus |

### Cross-references

<table>
<thead>
<tr>
<th>PIR</th>
<th>B27103, B27103.</th>
</tr>
</thead>
<tbody>
<tr>
<td>HSSP</td>
<td>POOS006; 7ATT[HSS/PDB]</td>
</tr>
</tbody>
</table>

### Comments

- **CATALYTIC ACTIVITY**: L-aspartate + 2-oxoglutarate = oxaloacetate + L-glutamate.
- **COFACTOR**: Pyridoxal phosphate.
- **SUBUNIT**: Homodimer.
- **SUBCELLULAR LOCATION**: Mitochondrial matrix.
- **MISCELLANEOUS**: In eukaryotes there are cytoplasmic, mitochondrial and chloroplastic isoforms.
- **SIMILARITY**: Belongs to the class I pyridoxal-phosphate-dependent aminotransferase family.

### References

1. PROTEIN SEQUENCE, TISSUE=Liver; MEDLINE=85289123; PubMed Kuramitsu S, Jouve F, Konno H, "Aspartate aminotransferase..."
Ensembl Gene Report

Gene
BRCA2 (HUGO ID)

Ensembl Gene ID
ENSG00000139618

Genomic Location
View gene in genomic location: 31775476 - 31776405 bp (1.0 Mb) on chromosome 13
This gene is located in sequence: AL145212.9.1.183687

Description
Breast cancer type 2 susceptibility protein. [Source: Uniprot/SWISSPROT;Acc:P61587]

RefSeq dna:
NM_000569 [Target %id: 99; Query %id: 98] [seq]
NP_000569 [Target %id: 99; Query %id: 99] [seq]

Uniprot/SWISSPROT:
BRCA2_HUMAN [Target %id: 98; Query %id: 99] [alan]

Prediction Method
The following GO terms have been mapped to this entry via UniProt:
GO:0000724 [regulation of cell cycle] TAS
GO:0003676 [nucleic acid binding] IBA
GO:003627 [single-stranded DNA binding] NAS
GO:003713 [transcription coactivator activity] NR
GO:003519 [protein binding] IBA
GO:002315 [extracellular space] TAS
GO:0005634 [nucleus] IBA
GO:0005221 [DNA repair] IBA
GO:0006326 [establishment and/or maintenance of chromatin architecture] TAS
GO:0006338 [chromatin remodeling] NAS
GO:0007030 [regulation of S phase of mitotic cell cycle] NAS
GO:0007083 [mitotic checkpoint] NAS
GO:0030141 [secretory granule] IBA
GO:0046449 [regulation of transcription] TAS

InterPro
IPR002093 [BRCA2 repeat - View other Ensembl genes with this domain]
IPR001887 [ATP/GTP-binding site motif A (P-loop) - View other Ensembl genes with this domain]

Protein Family
ENSP0000033848 : BREAST CANCER TYPE 2 SUSCEPTIBILITY
This cluster contains 1 Ensembl gene member(s)

Transcript Structure

www.ensembl.org/
Gene Page for mhkA

General Information
- **Gene Name**: mhkA
- **Synonym**: MHCK-A
- **Gene Product**: protein serine/threonine kinase, myosin heavy chain kinase, alpha kinase
- **Description**: myosin II heavy chain kinase that contains an N-terminal coiled-coil domain, a central MHCK domain, and a C-terminal domain with 7 WD-40 repeats; drives myosin II disassembly
- **Name Description**: mhk = Myosin Heavy chain Kinase
- **dictyBaseID**: DDB0216274

Chromosomal Coordinates
- Chromosome 6 coordinates: 15900 to 22490, Watson strand

Protein Information
- **Length**: 1,146 aa
- **Molecular Weight**: 128,963 Da
- **Protein Domain**: Predicted Protein Domains at GeneDB

Gene Ontology Annotations for mhkA
- **Molecular Function**: actin binding (MP), myosin II binding (IPI), ATP binding (IC), myosin-heavy-chain kinase activity (IDA), protein serine/threonine kinase activity (IDA), actin filament binding (IDA)
- **Biological Process**: autophosphorylation (IDA), protein amino acid phosphorylation (IDA), myosin II filament disassembly (IMP), myosin II filament disassembly (IDA)
- **Cellular Component**: cell cortex (IDA), cytoplasm (IDA)

Expression
- UCSD Expression Profile | BCM Expression Profile

Phenotype
- **Null**: increased myosin II assembly
- **Overexpression**: reduced growth in suspension, development arrests at mound stage, aberrant cytokinesis, reduced myosin II assembly

Associated Sequences
- **Curated Model**: DDB021
- **Gene Prediction From Sequencing Center**: DDB016
- **GenBank Genomic Fragment**: DDB000
- **GenBank mRNA**: DDB000
- **ESTs**: DDB0024936, DDB000
Outline

1. Introduction to the Gene Ontology
2. Gene Ontology annotations
3. Editing the Gene Ontology
4. Practical applications for the Gene Ontology
5. The Gene Ontology as one of many biological ontologies
How is GO maintained?

• Several full-time editors
• Requests from community
  – database curators, researchers, software developers
  – SourceForge tracker
• GO Consortium meetings for large changes
• Mailing lists
Ensuring Stability in a Dynamic Ontology

• Terms become obsolete when they are removed or redefined
• GO IDs are never deleted
• For each term, a comment is added to explains why the term is now obsolete
Why modify the GO

• GO reflects *current* knowledge of biology

• New organisms being added makes existing terms arrangements incorrect

• Not everything perfect from the outset
Example - parasites

• Original GO:

- cellular_component
  - cell
    - intracellular
      - intracellular organelle
        - intracellular membrane-bound organelle
          - nucleus
Example - parasites

- Annotation of *P. falciparum*
  - protozoan cellular parasite
  - intracellular infection (erythrocytes)
- Parasite proteins located in host nucleus
- What cellular component term to annotate to?
  - ‘nucleus’ refers to parasite nucleus when annotating parasite
Example - parasites

• Added new term ‘host’:

- cellular_component
  - extracellular region
    - extraorganismal space
      - host cell nucleus
Example - parasites

- parasite gene products located in parasite nucleus annotated here
- parasite gene products located in host nucleus annotated here
Requesting changes to GO - curator requests tracker

• Common changes suggested:
  – new term requests
  – reporting errors (typos, etc)
  – obsoletion/merge requests
  – add synonym
  – queries
  – term move (change parents)
The GO editorial office

• Primary responsibility to edit ontologies in response to community needs

• Also:
  – website
  – documentation
  – outreach
    • GO in other systems
    • new annotation groups
  – training
Outline

1. Introduction to the Gene Ontology
2. Gene Ontology Annotations
3. Editing the Gene Ontology
4. Practical applications for the Gene Ontology
5. The Gene Ontology as one of many biological ontologies
What can scientists do with GO?

• Access gene product functional information
• Find how much of a proteome is involved in a process/function/component in the cell
• Map GO terms and incorporate manual annotations into own databases
• Provide a link between biological knowledge and …
  • gene expression profiles
  • proteomics data
...analysis of high-throughput data according to GO

MicroArray data analysis

Bregje Wertheim at the Centre for Evolutionary Genomics,
Department of Biology, UCL and Eugene Schuster Group, EBI.
Selecting microarray subsets based on GO reveals drug target

Analysis of Functional Annotation – Downregulated Genes

GO terms

- 04110_cell cycle
- GO:0007067_mitosis
- GO:0000278_mitotic cell cycle
- GO:0000279_M phase
- PCNA
- E2F1
- CCNA2
- CDK2
- CDC6
- CACGTG_V$MYC_Q6
- V$E2F_Q6
- 00970_aminoacyl-tRNA biosynthesis
- GO:0005840_ribosome
- GO:0043039_tRNA aminoacylation
- GO:0006416_tRNA aminoacylation for protein translation
- 03050_proteasome

courtesy of Shabana Shabeer, Albert Einstein School of Medicine

Figure modified from http://en.wikipedia.org/wiki/Image:Microarray-schema.jpg
<table>
<thead>
<tr>
<th>Tools for gene expression/microarray analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>CLENCH</td>
</tr>
<tr>
<td>EASE</td>
</tr>
<tr>
<td>ermineJ</td>
</tr>
<tr>
<td>FuncAssociate</td>
</tr>
<tr>
<td>GARRAN</td>
</tr>
<tr>
<td>GFINDER: Genome Function</td>
</tr>
<tr>
<td>GoMiner and MatchMiner</td>
</tr>
<tr>
<td>GoSurfer</td>
</tr>
<tr>
<td>GOTM (Gene Ontology Tree Machine)</td>
</tr>
<tr>
<td>MAPPFinder</td>
</tr>
<tr>
<td>Onto-Compare</td>
</tr>
<tr>
<td>Onto-Express</td>
</tr>
<tr>
<td>Onto-Translate</td>
</tr>
<tr>
<td>Ontologizer</td>
</tr>
<tr>
<td>SeqExpress</td>
</tr>
<tr>
<td>THEA</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>
For monday.....

Install Blast2GO

www.blast2go.com
Outline

1. Introduction to the Gene Ontology
2. Gene Ontology Annotations
3. Editing the Gene Ontology
4. Practical applications for the Gene Ontology
5. The Gene Ontology as one of many biological ontologies
Beyond GO – Open Biomedical Ontologies

- Orthogonal to existing ontologies to facilitate combinatorial approaches
  - Share unique identifier space
  - Include definitions

- Anatomies
- Cell Types
- Sequence Attributes
- Temporal Attributes
- Phenotypes
- Diseases
- More….

http://obo.sourceforge.net
Sequence Ontology

http://song.sourceforge.net
• Ontology of ‘small molecular entities’

http://www.ebi.ac.uk/chebi
Review

1. What is an ontology?
2. Why do we need ontologies?
3. What are the three gene ontologies?
4. Give an example of a is-a relationship in the Cellular Component ontology
5. Give an example of the part-of relationship in the Cellular Component ontology.
6. …in the Molecular Function ontology.
7. …in the Biological process ontology
8. What two things must every annotation have?
9. What are evidence codes?
10. What is the least confident evidence code?