

Ontologies for biology

Karen Eilbeck

Programming for Biology

CSHL 2017

What we will cover

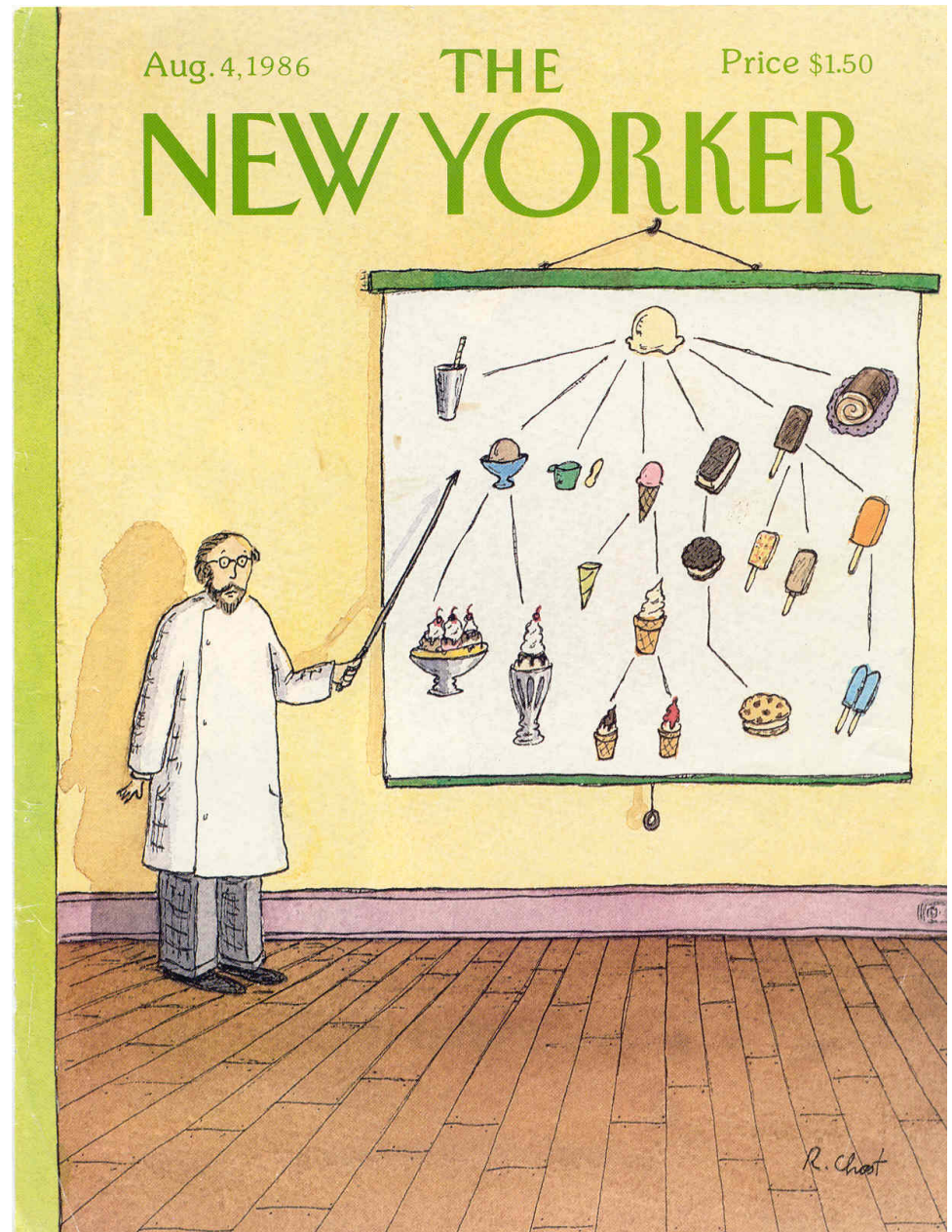
- Introduction to ontologies
- Gene Ontology case study
 - Structure
 - Annotation
 - Usage
- Other biological ontology applications
- Upper ontologies and management issues

Building ontologies for the genomics *big data* challenge

- We have more data than we can handle by eye
- Data needs to be described in a rigorous fashion to render it computable
- Functional annotation using the [Gene Ontology](#) (GO)
- Structural annotation using the [Sequence Ontology](#) (SO)
- Human Phenotype Ontology (HPO)



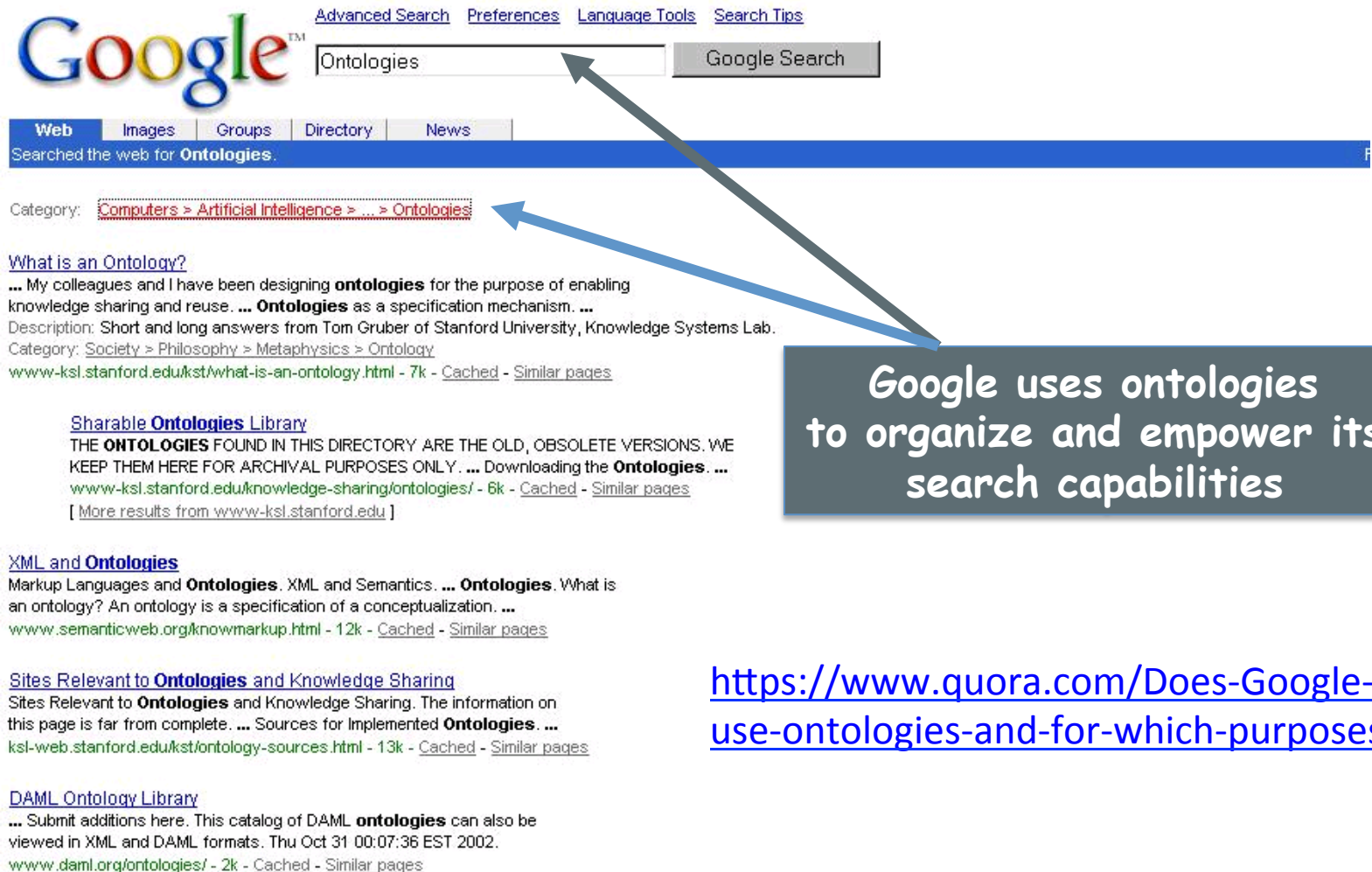
- Ice Cream Ontology



Ontologies are:

- Semantic models for domains of reality
 - Tool for communication
 - Explain observations
 - Predict new relations
 - Provide framework for data integration

Ontologies in everyday life



The screenshot shows a Google search interface. The search bar contains the word "Ontologies". Below the search bar, the results are categorized under "Computers > Artificial Intelligence > ... > Ontologies". A blue arrow points from the search bar to the category path, and a grey arrow points from the category path to a text box on the right.

Google uses ontologies to organize and empower its search capabilities

[Advanced Search](#) [Preferences](#) [Language Tools](#) [Search Tips](#)

Google™

Ontologies Google Search

Web Images Groups Directory News

Searched the web for **Ontologies**.

Category: [Computers > Artificial Intelligence > ... > Ontologies](#)

[What is an Ontology?](#)
... My colleagues and I have been designing **ontologies** for the purpose of enabling knowledge sharing and reuse. ... **Ontologies** as a specification mechanism. ...
Description: Short and long answers from Tom Gruber of Stanford University, Knowledge Systems Lab.
Category: [Society > Philosophy > Metaphysics > Ontology](#)
www-ksl.stanford.edu/ksst/what-is-an-ontology.html - 7k - [Cached](#) - [Similar pages](#)

[Sharable Ontologies Library](#)
THE **ONTOLOGIES** FOUND IN THIS DIRECTORY ARE THE OLD, OBSOLETE VERSIONS. WE KEEP THEM HERE FOR ARCHIVAL PURPOSES ONLY. ... Downloading the **Ontologies**. ...
www-ksl.stanford.edu/knowledge-sharing/ontologies/ - 6k - [Cached](#) - [Similar pages](#)
[[More results from www-ksl.stanford.edu](#)]

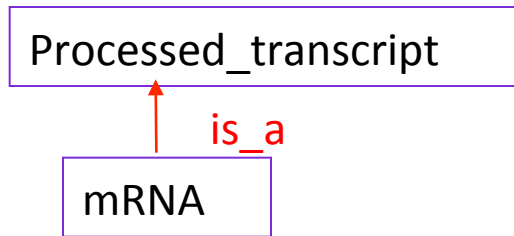
[XML and Ontologies](#)
Markup Languages and **Ontologies**. XML and Semantics. ... **Ontologies**. What is an ontology? An ontology is a specification of a conceptualization. ...
www.semanticweb.org/knowmarkup.html - 12k - [Cached](#) - [Similar pages](#)

[Sites Relevant to Ontologies and Knowledge Sharing](#)
Sites Relevant to **Ontologies** and Knowledge Sharing. The information on this page is far from complete. ... Sources for Implemented **Ontologies**. ...
ksl-web.stanford.edu/ksst/ontology-sources.html - 13k - [Cached](#) - [Similar pages](#)

[DAML Ontology Library](#)
... Submit additions here. This catalog of DAML **ontologies** can also be viewed in XML and DAML formats. Thu Oct 31 00:07:36 EST 2002.
www.daml.org/ontologies/ - 2k - [Cached](#) - [Similar pages](#)

<https://www.quora.com/Does-Google-use-ontologies-and-for-which-purposes>

The **is_a** relationship



hypernym

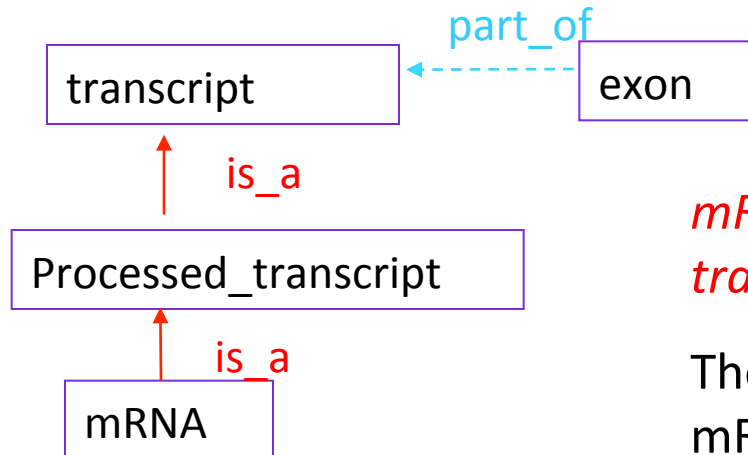


hyponym

- Subsumption hierarchy
- General to specific categorization
- Define terms with regard to **is_a** (An A is a B that C's.)

The *is_a* relationship

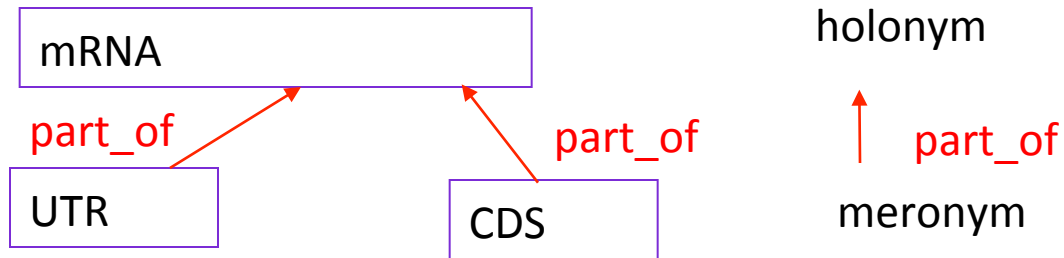
- The *is_a* relation is like inheritance.
- Children terms inherit the properties and relationships of the parent term.



mRNA inherits the attributes of *transcript*

Therefore *exons* can be parts of mRNA

The **part_of** relationship



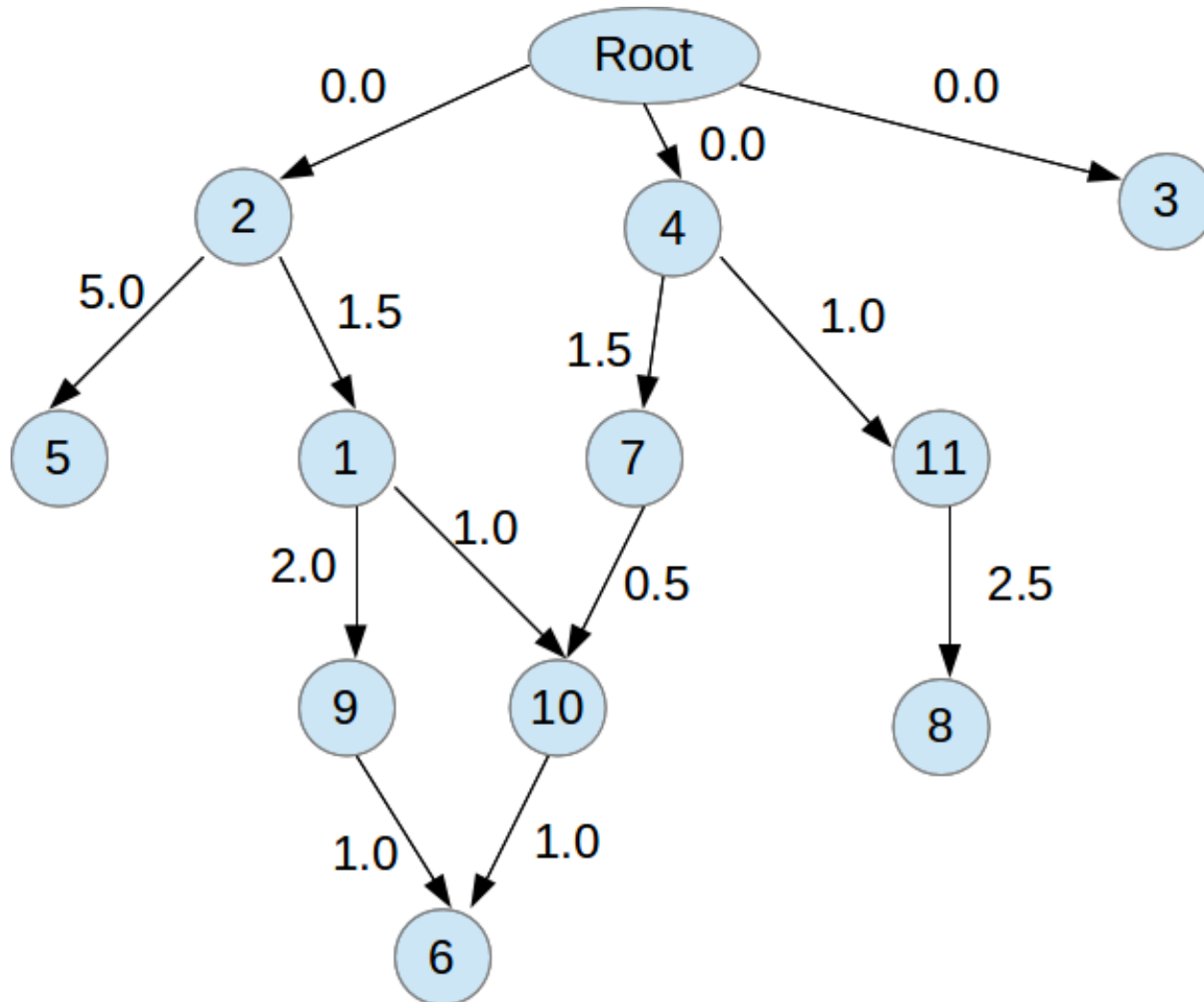
Meronomy

A dissection of a whole into constituent parts

The **part_of** relationship

- The rules of being a part:
 - Nothing is a part of itself
 - If **A** is a part of **B** then the **B** is not a part of **A**
 - If **A** is a part of **B** and **B** is a part of **C** then **A** is a part of **C**
 - The relationship is **asymmetrical** and **transitive**
- Parts do not inherit properties from the whole

Ontologies are often directed acyclic graphs DAG



Bio-ontologies

- Annotation of databases
- Semantically aware searching
- Inference and reasoning
 - Find inconsistencies in data
- Understanding high-throughput data
 - Over representation analysis
- Semantic similarity analysis

A biomedical ontology organizes data

- Classify the kinds of things we are interested in (domain).
 - Taxonomy
- Categorize the interactions between the things – define other relations.
- Ontologies are applied to the raw data via curation process.
- The data can then be traversed and manipulated computationally using the relationships.
- We have conquered the *big data*.

How do we keep on top of all the data?

- Biology papers are indexed by medline and live in a db called pubmed.
- <http://www.ncbi.nlm.nih.gov/pubmed/>
- You can search pubmed by names and terms.
- Behind pubmed is a terminology called MeSH for medical sub headings.
- It organizes synonyms, etc. to make the searches more efficient.
- This is still not enough, but it is a start.

Millions of abstracts are indexed.

Imagine if someone read the papers for you

- Gene Ontology functional annotation
- <http://www.geneontology.org>
- A massive multinational project, including many model organism database groups.
- An ontology of 30 000+ terms, organized into a hierarchy of specificity.
- Skilled curators who are paid to read articles and annotate the genes with the **function, process and location**.

Gene Ontology

- 1998
- The problem: a lot of sequence databases. No common vocabulary to talk about genes. No way to search sequences databases other than by similarity.
- The solution: describe function, process and location, directed acyclic graph, unify vocabulary, provide structure to aid annotation and querying



- Describes the location, function and process of a gene product (protein or functional RNA).
- MODs employ curators to transfer knowledge from the literature and bioinformatics experiments to the genome.
- 28,000 terms, 44,545,253 annotations
386,268 genes (52 databases)

How does GO help manage big data?

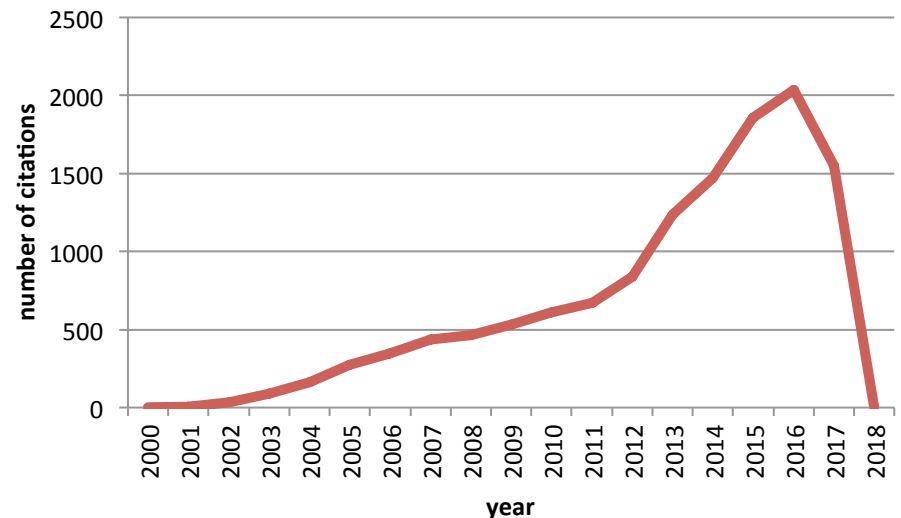
- Unified the descriptive vocabulary between the model organism communities.
- Structured the semantic nature of what genes do.
- Enables querying: It allows sequence databases to be queried on something other than sequence similarity.
- Downstream analyses use the GO annotations across the board in all aspects of biology.

Find transcription factors in multiple databases

- TF have diverse sequence structures. There are many ways to bind DNA. You can't do a similarity search with one kind of TF and find the rest.
- GO allows you to search a sequence database on semantics rather than sequence. So if the TF are annotated with GO, you could find them easily across multiple organism databases.

GO is useful for processing results of experiments.

- What kinds of protein are enriched in my microarray experiment?
- How does my newly sequenced boutique genome compare to a similar established model organism?
- Increasingly cited:



Structure

biological_process

cellular_component

molecular_function

biological_process

- ⊕ ← I behavior
- ⊕ ← I biological adhesion
- ⊕ ← I biological phase
- ⊕ ← I biological regulation
- ⊕ ← I carbohydrate utilization
- ⊕ ← I carbon utilization
- ⊕ ← I cell aggregation
- ⊕ ← I cell killing
- ⊕ ← I cell proliferation
- ⊕ ← I cellular component organization or bioger
- ⊕ ← I cellular process
- ⊕ ← I detoxification
- ⊕ ← I developmental process
- ⊕ ← I growth
- ⊕ ← I immune system process
- ⊕ ← I localization
- ⊕ ← I locomotion
- ⊕ ← I metabolic process
- ⊕ ← I multi-organism process
- ⊕ ← I multicellular organismal process
- ⊕ ← R negative regulation of biological process
- ⊕ ← I nitrogen utilization
- ⊕ ← I phosphorus utilization
- ⊕ ← I pigmentation
- ⊕ ← R positive regulation of biological process
- ⊕ ← I presynaptic process involved in chemical synaptic transmiss
- ⊕ ← R regulation of biological process
- ⊕ ← I reproduction
- ⊕ ← I reproductive process
- ⊕ ← I response to stimulus
- ⊕ ← I rhythmic process
- ⊕ ← I signaling
- ⊕ ← I sulfur utilization

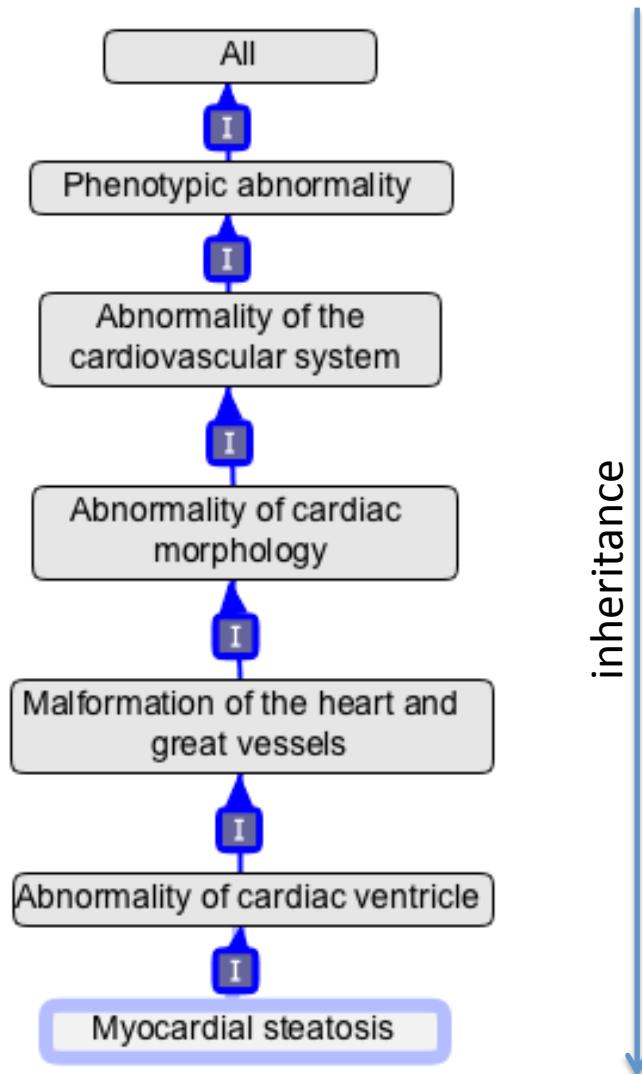
cellular_component

- ⊕ ← I cell
- ⊕ ← I cell junction
- ⊕ ← I cell part
- ⊕ ← I extracellular region
- ⊕ ← I extracellular region part
- ⊕ ← I macromolecular complex
- ⊕ ← I membrane
- ⊕ ← I membrane part
- ⊕ ← I membrane-enclosed lumen
- ⊕ ← I mitochondrion-associated adheren:
- ⊕ ← I nucleoid
- ⊕ ← I organelle
- ⊕ ← I organelle part
- ⊕ ← I other organism
- ⊕ ← I other organism part
- ⊕ ← I supramolecular complex
- ⊕ ← I symplast
- ⊕ ← I synapse
- ⊕ ← I synapse part
- ⊕ ← I viral occlusion body
- ⊕ ← I virion
- ⊕ ← I virion part

molecular_function

- ⊕ ← I antioxidant activity
- ⊕ ← I binding
- ⊕ ← I catalytic activity
- ⊕ ← I hijacked molecular function
- ⊕ ← I molecular carrier activity
- ⊕ ← I molecular function regulator
- ⊕ ← I molecular transducer activity
- ⊕ ← I nutrient reservoir activity
- ⊕ ← I protein tag
- ⊕ ← I signal transducer activity
- ⊕ ← I structural molecule activity
- ⊕ ← I toxin activity
- ⊕ ← I transcription regulator activity
- ⊕ ← I translation regulator activity
- ⊕ ← I transporter activity

True path rule



Genes annotated to myocardial steatosis is also annotated by transitivity to each parent node in the ontology

Increases flexibility when searching and making inferences about genes

Annotation

- <http://www.geneontology.org/page/download-annotations> The files
- <http://www.geneontology.org/page/go-annotation-file-formats> Understanding the columns
- [http://wiki.geneontology.org/index.php/Introduction to the GO Consortium annotation set](http://wiki.geneontology.org/index.php/Introduction_to_the_GO_Consortium_annotation_set) more info

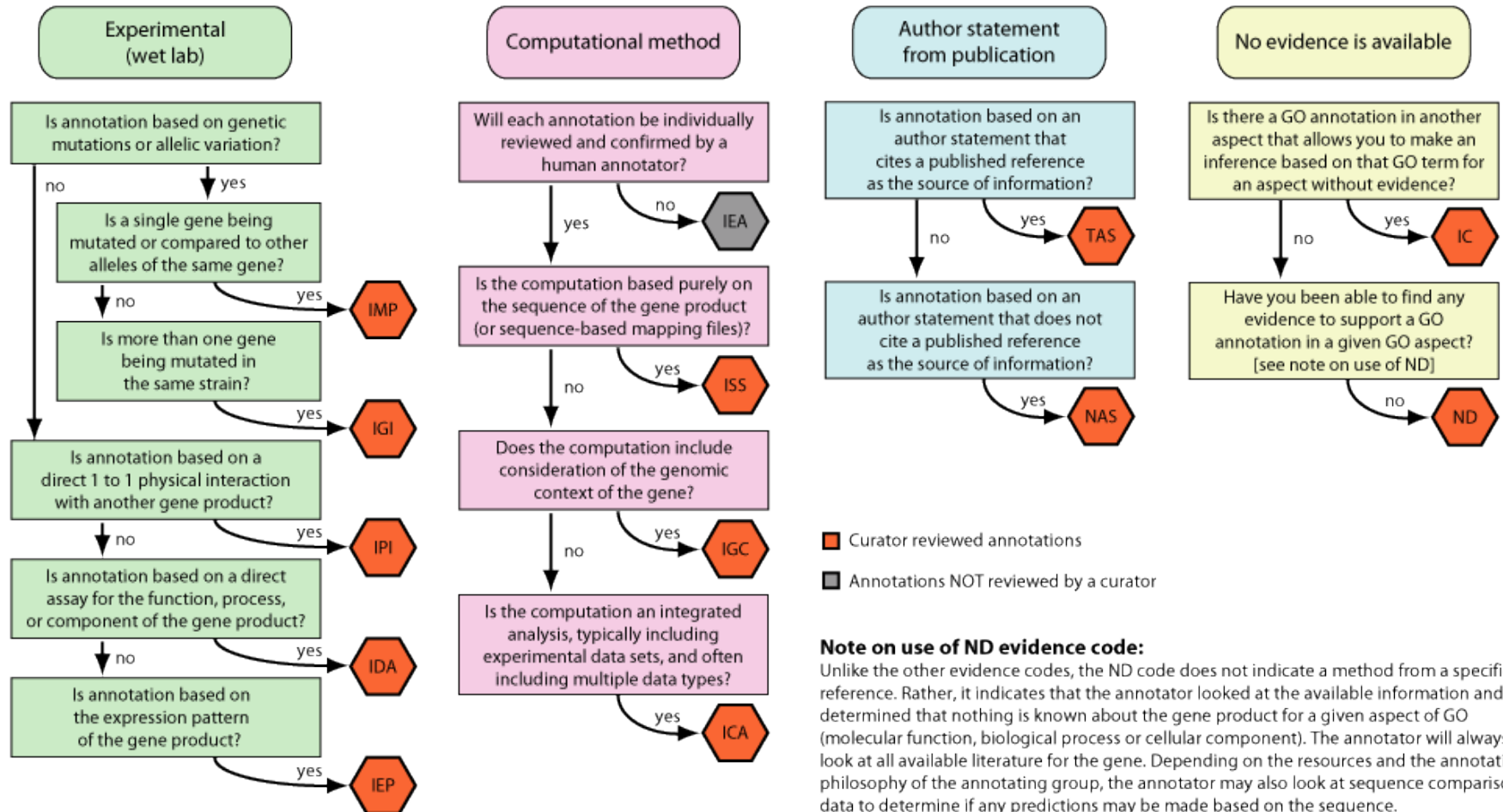
Annotation - simple

- GeneID – GO term – pubmedID – evidence code
- Annotate to the deepest level of knowledge
- Difference between un-annotated and uncharacterized – when you have tried to find function/process/location and the information is not out there, you annotate the gene to the root node.
- Qualifiers – contributes-to, co-localized with and **NOT**. Be careful!

Evidence

GO Evidence Code Decision Tree

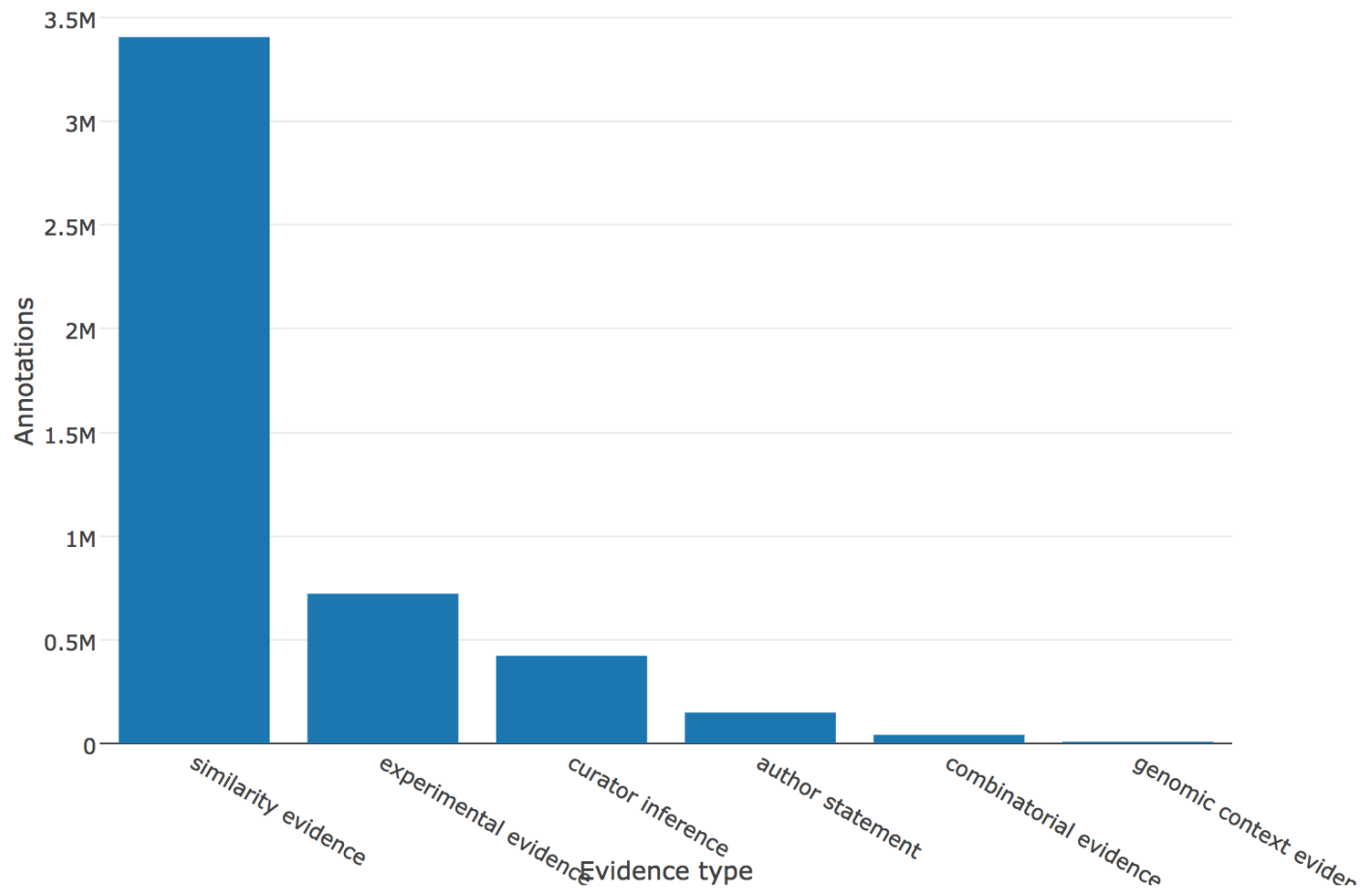
What type of evidence is the annotation based on?



The kinds of evidence:

- Experimental from literature (**direct**)
 - Expression pattern
 - Physical interaction
- Computational evidence (**inferred**)
 - Pairwise alignment
 - Sequence orthology
- Curator/author assertions (**indirect**)

Annotations by evidence



Usage

- How is this ontology used?
- High through-put studies that produced lists of genes micro-array and RNA-seq
- Functional categorization – what does this newly sequenced genome do?

What does the GO give us?

- For every gene in a genome we can get an understanding of what the gene is doing.
- If the gene has not been studied, we can infer from similar genes that have been.
- We can traverse a genome based on an understanding of function. We can use this to understand result of experiments that produce collections of genes under certain circumstances.
 - All of these genes have something to do with cell cycle control etc

Over representation analysis

- GO has annotated millions of gene products
- RNA-seq detects expression of these genes and produce lists of differentially expressed genes.
- Does a GO term show up more than you would expect? (Your usual observed/expected kind of analysis)

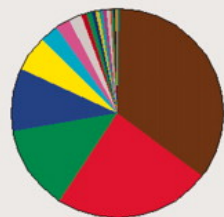
Quick Yeast example

- 221 yeast genes labeled with sporulation (out of 6000)
- In a microarray experiment 100 genes are up-regulated compared to control
- You would expect 3-4 to be annotated to sporulation by chance
- If you found 40, that would be over representation

Over representation tools

- <http://david.abcc.ncifcrf.gov/> now provides a comprehensive set of functional annotation tools for investigators to understand biological meaning behind large list of genes
- Go term Finder
<http://go.princeton.edu/cgi-bin/LAGO>

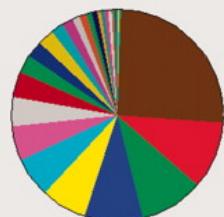
Functional Categorization for : GO Cellular Component



cellular_component unknown:	35.3% (raw value = 10904)
membrane:	23.8% (raw value = 7348)
plastid:	12.9% (raw value = 3990)
mitochondrion:	9.8% (raw value = 3034)
nucleus:	5.5% (raw value = 1701)
intracellular:	2.9% (raw value = 894)
cytoplasm:	2.1% (raw value = 644)
ribosome:	1.9% (raw value = 586)
thylakoid:	0.7% (raw value = 227)
cell:	0.7% (raw value = 226)
cytoskeleton:	0.6% (raw value = 197)
unlocalized:	0.6% (raw value = 175)
plasma membrane:	0.5% (raw value = 161)
extracellular:	0.5% (raw value = 155)
cell wall:	0.5% (raw value = 143)
cytosol:	0.3% (raw value = 105)
extracellular matrix:	0.3% (raw value = 80)
endoplasmic reticulum:	0.3% (raw value = 80)
nucleoplasm:	0.2% (raw value = 61)
Golgi apparatus:	0.2% (raw value = 47)
peroxisome:	0.1% (raw value = 31)
vacuole:	0.1% (raw value = 27)
other:	0% (raw value = 14)
nuclear membrane:	0% (raw value = 9)
external encapsulating structure:	0% (raw value = 9)
nucleolus:	0% (raw value = 8)
lysosome:	0% (raw value = 7)
endosome:	0% (raw value = 5)
extracellular space:	0% (raw value = 1)

a

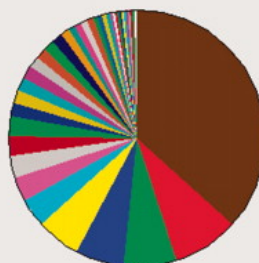
Functional Categorization for : GO Molecular Function



molecular_function unknown:	26% (raw value = 10319)
transferase activity:	10.1% (raw value = 4023)
catalytic activity:	9.8% (raw value = 3880)
hydrolase activity:	8.9% (raw value = 3533)
nucleotide binding:	6.9% (raw value = 2719)
transporter activity:	6.2% (raw value = 2449)
DNA binding:	5.3% (raw value = 2109)
transcription factor activity:	4.4% (raw value = 1738)
binding:	3.8% (raw value = 1490)
kinase activity:	2.6% (raw value = 1048)
protein binding:	2.6% (raw value = 1019)
nucleic acid binding:	2.5% (raw value = 992)
RNA binding:	1.9% (raw value = 753)
structural molecule activity:	1.6% (raw value = 622)
other:	1.1% (raw value = 454)
receptor activity:	1% (raw value = 407)
nuclease activity:	0.8% (raw value = 333)
chaperone activity:	0.7% (raw value = 284)
enzyme regulator activity:	0.6% (raw value = 249)
oxygen binding:	0.6% (raw value = 229)
translation factor activity, nucleic acid binding:	0.5% (raw value = 203)
transcription regulator activity:	0.5% (raw value = 203)
lipid binding:	0.4% (raw value = 155)
signal transducer activity:	0.4% (raw value = 149)
carbohydrate binding:	0.3% (raw value = 116)
motor activity:	0.2% (raw value = 91)
receptor binding:	0.2% (raw value = 68)
chromatin binding:	0.1% (raw value = 30)

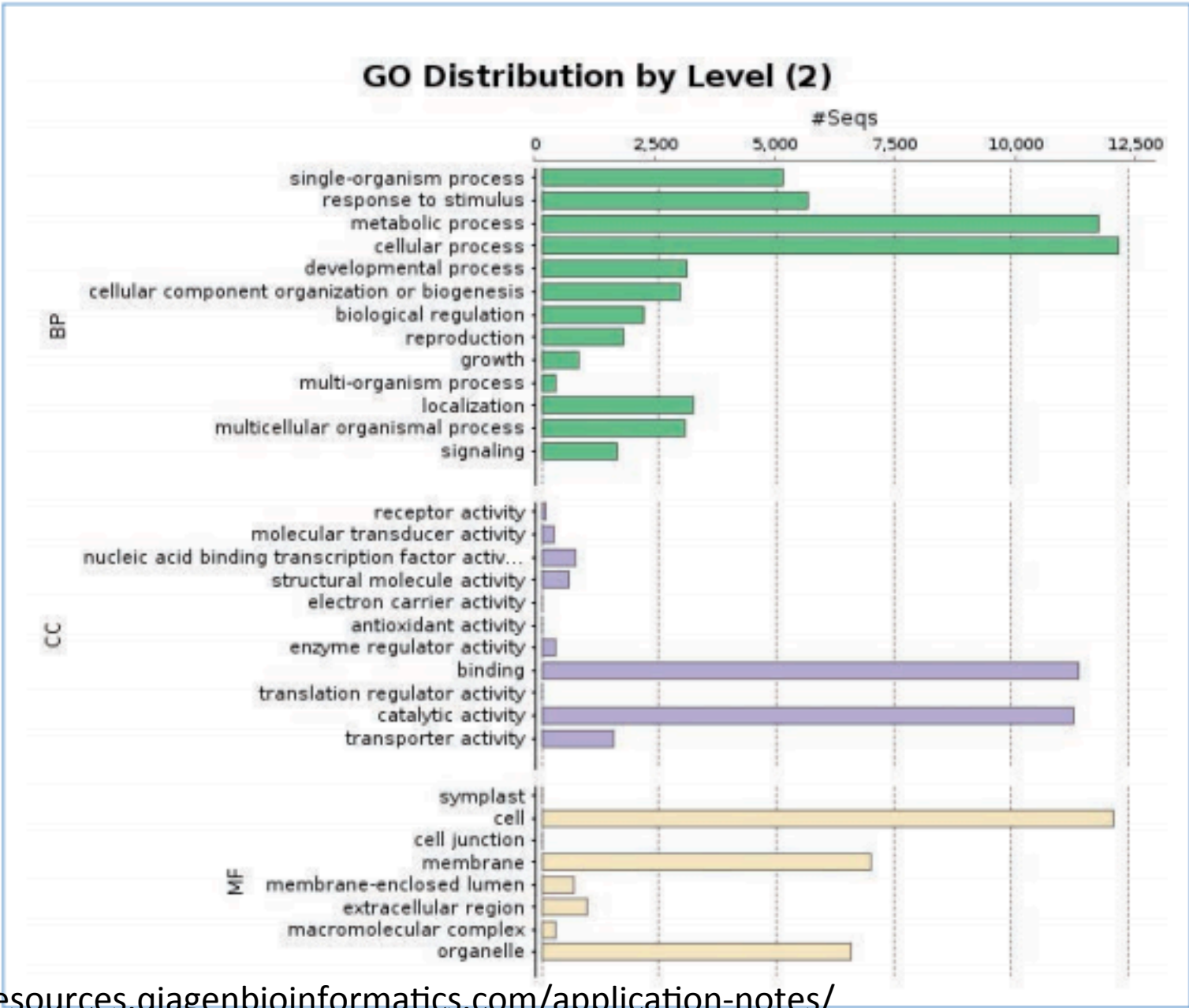
b

Functional Categorization for : GO Biological Process



biological_process unknown:	36.6% (raw value = 12814)
transport:	8.5% (raw value = 2964)
metabolism:	6.6% (raw value = 2305)
transcription:	6.1% (raw value = 2134)
protein metabolism:	5.7% (raw value = 1990)
protein modification:	3.7% (raw value = 1291)
electron transport:	3% (raw value = 1043)
protein biosynthesis:	2.7% (raw value = 929)
signal transduction:	2.6% (raw value = 901)
carbohydrate metabolism:	2.4% (raw value = 840)
DNA metabolism:	1.8% (raw value = 625)
cell organization and biogenesis:	1.7% (raw value = 612)
cell cycle:	1.7% (raw value = 604)
response to abiotic stimulus:	1.7% (raw value = 588)
response to stress:	1.4% (raw value = 482)
nucleobase, nucleoside, nucleotide and nucleic acid metabolism:	
biosynthesis:	1.3% (raw value = 461)
response to endogenous stimulus:	1.3% (raw value = 452)
energy pathways:	1% (raw value = 355)
physiological process:	1% (raw value = 339)
development:	0.9% (raw value = 325)
lipid metabolism:	0.9% (raw value = 303)
response to biotic stimulus:	0.9% (raw value = 302)
cell death:	0.5% (raw value = 188)
morphogenesis:	0.5% (raw value = 180)
secondary metabolism:	0.5% (raw value = 163)
other:	0.4% (raw value = 153)
cell growth and/or maintenance:	0.4% (raw value = 128)
photosynthesis:	0.3% (raw value = 122)
cellular process:	0.3% (raw value = 107)
reproduction:	0.3% (raw value = 91)
embryonic development:	0.3% (raw value = 88)
flower development:	0.2% (raw value = 84)
regulation of gene expression, epigenetic:	0.2% (raw value = 78)
cell communication:	0.2% (raw value = 73)
cell differentiation:	0.2% (raw value = 71)
catabolism:	0.2% (raw value = 69)
post-embryonic development:	0.2% (raw value = 68)
amino acid and derivative metabolism:	0.1% (raw value = 43)
cell-cell signaling:	0.1% (raw value = 33)
tropism:	0.1% (raw value = 31)
growth:	0.1% (raw value = 27)
behavior:	0.1% (raw value = 22)
post-pollination:	0.1% (raw value = 22)
response to extracellular stimulus:	0.1% (raw value = 22)
response to external stimulus:	0% (raw value = 14)
ripening:	0% (raw value = 3)
cell homeostasis:	0% (raw value = 3)
pollen-pistil interaction:	0% (raw value = 3)
abscission:	0% (raw value = 1)
cell growth:	0% (raw value = 1)

c



How do I make high level summaries like that?

- <http://www.geneontology.org/page/go-slim-and-subset-guide>
- Slim ontology – hand pruned tope level terms.
- Propagate annotations up the tree to the new leaf node.

Resources

- Use and misuse of the gene ontology annotations S Rhee et al.

<http://www.nature.com/nrg/journal/v9/n7/full/nrg2363.html>

- Ten Quick Tips for Using the Gene Ontology

<http://www.nature.com/nrg/journal/v9/n7/full/nrg2363.html> Judith Blake

2 other bio ontologies



- The SO describes the parts of a genome and how they relate to each other in topological space and other dimensions such as regulatory space.
- It includes terms and relations to describe not only gene models but other genomic phenomena such as transposons and repeats.
- It also describes variations and the consequences of variation.
- About 2000 terms

SO mission statement

To provide the vocabulary and relations for the annotation, validation and querying of the features and properties of biological sequence.

In other words describe genome assemblies and their annotations

SO is used for

- Genome annotation
 - Typing features in Model Organism databases (Chado)
 - Typing variant data in re-sequencing projects
 - And file formats - GFF3, GVF
- Genome browsing - Gbrowse
- Protein feature annotation - biosapiens
- Natural Language Processing – RSC
- ICCG, ICGC, UCSC browser, EBI, NCBI ClinVar, DBVar

What do the sequence annotators want from SO

- To annotate the sequence of the assembly - contigs, scaffolds, gaps, etc.
- To annotate the experimental evidence for gene annotations- blast hits, repeats.
- To annotate known and postulated biological features - exons, transposons.
- Now they also want to annotate variants.

How does SO help manage big data?

- It has standardized the vocabulary used to define the parts of genomes.
- It is used by the MODs to annotate features onto their genomes.
- It is used by EBI, personal genome companies and institutes to manage variant data.
- The contents of a genome are now more easily analyzed by **ontology aware software** tools.

Organisms with SO based genome annotations:

Anopheles spp., *Antonospora locustae*, *Arabidopsis thaliana*, Ascomycota, *Avena* spp., Basidiomycota, *Bos taurus*, *Brachionus plicatilis*, *Caenorhabditis* spp., *Ciano intestinalis*, *Cryptosporidium* sp., *Culex quinquefasciatus*, *Danio rerio*, *Daphnia pulex*, *Dictyostelium discoideum*, *Drosophila* spp., *Emiliana huxleyi*, *Franciscella tularensis*, *Giardia* spp., *Glycine max*, *Homo sapiens*, *Ixodes scapularis*, *Lotus japonicus*, *Medicago truncatula*, Microsporidia, *Mycobacterium tuberculosis*, *Oryza* spp., *Paramecium tetraurelia*, *Pediculus humanus*, *Plasmodium* spp., *Populus trichocarpa*, *Pseudomonas* spp., *Ricinus communis*, *Rattus norvegicus*, Rubiaceae, *Saccharomyces* spp., *Schizosaccharomyces pombe*, *Schmidtea mediterranea*, Solanaceae, *Sorghum bicolor*, *Strongylocentrotus purpuratus*, *Tetrahymena thermophila*, *Theileria* spp., *Toxoplasma* spp., *Tribolium castaneum*, *Trichomonas* spp., Triticeae, *Trypanosoma brucei*, *Vitis vinifera*, *Zea mays*, Zygomycota, the influenza virus

... and many many more...

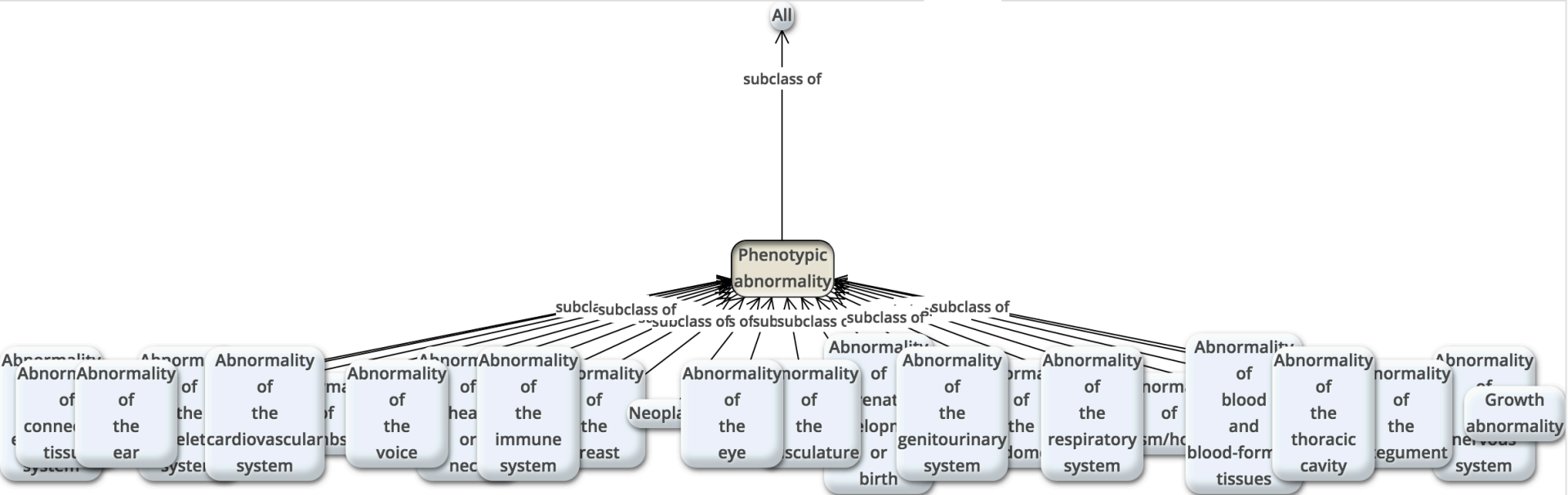
SO facilitates:

- Data sharing
- Comparative genomics - all data uses same language.
- Genomics tool development
- Quality control - traversing the relationships implied by SO is used to validate annotation (syntactic, semantic and topological)
- Annotation exploration - how are annotations evolving?

Human Phenotype Ontology

- <http://www.human-phenotype-ontology.org>
- The Human Phenotype Ontology (HPO) aims to provide a standardized vocabulary of phenotypic abnormalities encountered in human disease.
- Browse
<http://compbio.charite.de/phenexplorer/>

Organized by body system



You can also view the structure here: <https://bioportal.bioontology.org/ontologies/HP>

Example of phenotype and annotation

- Abnormality of the face
 - Abnormal facial expression
 - Abnormal facial shape
 - Abnormality of facial soft tissue
 - Abnormality of the chin
 - Abnormality of the forehead
 - Abnormality of the midface
 - Abnormality of the mouth
 - Abnormality of the nose
 - Abnormality of the orbital region
 - Abnormal morphology of the ocular adnexa
 - Abnormality of the common tendinous ring
 - Abnormality of the conjunctiva
 - Abnormality of the extraocular muscles
 - Abnormality of the eyebrow
 - Abnormality of the eyelid
 - Abnormality of levator palpebrae superioris
 - Abnormality of the eyelashes
 - Abnormality of the palpebral fissures
 - Abnormal shape of the palpebral fissure
 - Abnormal size of the palpebral fissures
 - Slanting of the palpebral fissure
 - Downslanted palpebral fissures**
 - Upslanted palpebral fissure

Gene	Associated diseases
TPM2 (7169)	MYOPATHY 4 (OMIM:609285), CONGENITAL FIBER CONGENITAL, WITH FIBER-TYPE DI... (OMIM:30058 TYPE 2B (OMIM:601680)
GABRD (2563)	GENERALIZED EPILEPSY WITH FEBRILE SEIZUR... (C AWAK... (OMIM:607628), EPILEPSY, MYOCLONIC JU FEBRILE SEIZUR... (OMIM:604233), EPILEPSY, IDIOP WITH FEBRILE SEIZUR... (OMIM:616172), MYOCLON GENERALIZED EPILEPSY WITH FEBRILE SEIZUR... (C SUSCEP... (OMIM:613060)
MKKS (8195)	BARDET-BIEDL SYNDROME 4 (OMIM:615982), BAR (OMIM:615988), BARDET-BIEDL SYNDROME (ORP SYNDROME 6 (OMIM:605231), BARDET-BIEDL SYN MCKUSICK-KAUFMAN SYNDROME (OMIM:236700), (OMIM:615990), BARDET-BIEDL SYNDROME 2 (OM SYNDROME 12 (OMIM:615989), BARDET-BIEDL SYI
TSR2 (90121)	BLACKFAN-DIAMOND ANEMIA (ORPHANET:124), D ANEMIA 6 (OMIM:612561), DIAMOND-BLACKFAN A ANEMIA 8 (OMIM:612563), DIAMOND-BLACKFAN A ANEMIA 12 (OMIM:615550), DIAMOND-BLACKFAN DIAMOND-BLACKFAN ANEMIA 13 (OMIM:615909),

Disease id	Disease name	Associated genes
OMIM:602501	MEGALENCEPHALY-CAPILLARY MALFORMATION-POLYMICROGYRIA SYNDROME	PIK3CA (5290)
OMIM:208050	ARTERIAL TORTUOSITY SYNDROME	SLC2A10 (81031)
OMIM:616580	AU-KLINE SYNDROME	
OMIM:149730	LACRIMOauriculodentodigital syndrome	FGFR3 (2261), FGFR2 (2263), FGF10 (2255)
OMIM:227260	FACIAL ECTODERMAL DYSPLASIA	TWIST2 (117581)
OMIM:180870	RUVALCABA SYNDROME	
OMIM:602471	SHORT STATURE, AUDITORY CANAL ATRESIA, MANDIBULAR HYPOPLASIA, AND SKELETAL ABNORMALITIES	GSC (145258)
OMIM:614230	CHROMOSOME 8Q21.11 DELETION SYNDROME	

Phenomizer

Menu ▾ Support the Phenomizer. Help. The Phenomizer

Features. Diseases. Ontology.

myopathy

HPO id.	Feature.
HP:0003198	Myopathy
HP:0008978	Necrotizing myopathy
HP:0011664	Non-compaction cardiomyopathy
HP:0005152	Oncocytic cardiomyopathy
HP:0001723	Restrictive cardiomyopathy
HP:0011663	Right ventricular cardiomyopathy
HP:0009054	Scapuloperoneal myopathy
HP:0003756	Skeletal myopathy
HP:0011665	Takotsubo cardiomyopathy

Page 2 of 2

Features 16 - 24 of 24

Patient's Features. **Diagnosis.**

Algorithm: resnik (Unsymmetric). 4 Features.

<input type="checkbox"/>	p-value. ▲	Disease Id.	Disease name.	Genes.
<input type="checkbox"/>	0.8371	OMIM:13...	#131300 CAMURATI-ENGELMANN DISEASE; ...	TGFB1
<input type="checkbox"/>	0.8371	OMIM:61...	#614557 EHLERS-DANLOS SYNDROME WIT...	FKBP14
<input type="checkbox"/>	0.8371	OMIM:10...	ANAL SPHINCTER MYOPATHY, INTERNAL	
<input type="checkbox"/>	1.0000	OMIM:19...	#194050 WILLIAMS-BEUREN SYNDROME; W...	
<input type="checkbox"/>	1.0000	OMIM:60...	#606054 PROPIONIC ACIDEMIA;;PROPIONY...	PCCB, PCCA
<input type="checkbox"/>	1.0000	OMIM:25...	SCHWARTZ-JAMPEL SYNDROME, TYPE 1	HSPG2
<input type="checkbox"/>	1.0000	OMIM:30...	#307030 HYPERGLYCEROLEMIA;;GLYCERO...	GK
<input type="checkbox"/>	1.0000	OMIM:22...	FANCONI-BICKEL SYNDROME	SLC2A2
<input type="checkbox"/>	1.0000	OMIM:21...	PITUITARY ADENOMA, ACTH-SECRETING	AIP
<input type="checkbox"/>	1.0000	OMIM:61...	#613327 LIPODYSTROPHY, CONGENITAL G...	PTRF
<input type="checkbox"/>	1.0000	OMIM:25...	#252500 MUCOLIPIDOSIS II ALPHA/BETA;;M...	GNPTAB
<input type="checkbox"/>	1.0000	OMIM:17...	#176270 PRADER-WILLI SYNDROME; PWS;;P...	SNRPN, NDN
<input type="checkbox"/>	1.0000	OMIM:18...	#182212 SHPRINTZEN-GOLDBERG CRANIO...	SKI
<input type="checkbox"/>	1.0000	OMIM:21...	#212065 CONGENITAL DISORDER OF GLYC...	PMM2
<input type="checkbox"/>	1.0000	OMIM:12...	ARTHROGRYPOSIS, DISTAL, TYPE 9	FBN2
<input type="checkbox"/>	1.0000	OMIM:01...	#010001 BONE FRAGILITY WITH CONTRAS...	SLC30A1

Page 1 of 248

Improve Differential Diagnosis. [Download Results.](#)

scoliosis, osteopenia, myopathy

Genes to disease annotation

Phenotypes to disease annotation

Upper ontologies and management

Places to find biomedical ontologies

- The OBO foundry www.obofoundry.org
- The NCBO www.bioontology.org

The goal of the National Center for Biomedical Ontology is to support biomedical researchers in their knowledge-intensive work, by providing online tools and a web portal enabling them to access, review, and integrate disparate ontological resources in all aspects of biomedical investigation and clinical practice. A major focus of our work involves the use of biomedical ontologies to aid in the management and analysis of data derived from complex experiments.

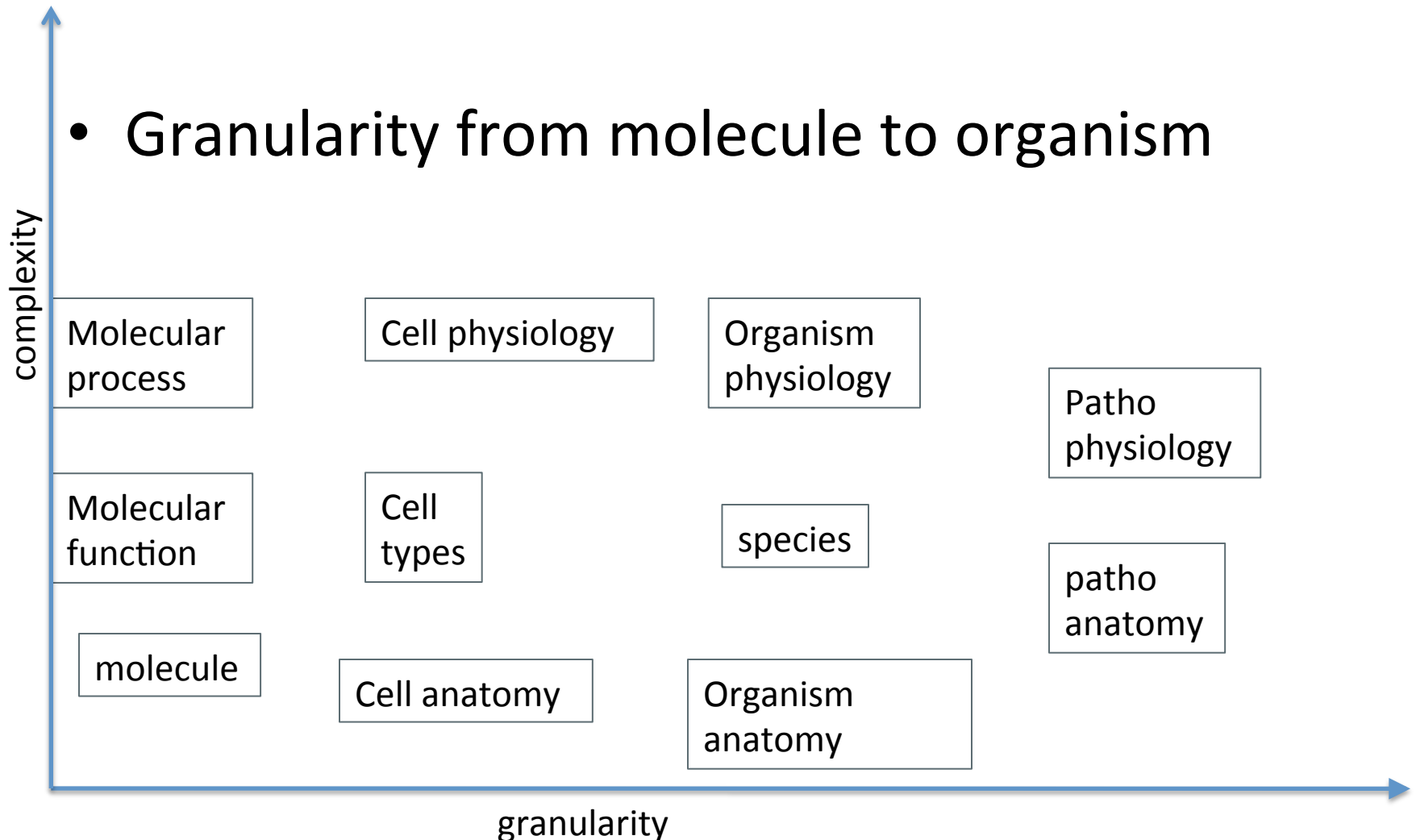


OBO Foundry

- Open
- Common syntax
- Unique identifier space
- Versioning control
- orthogonality
- Definitions
- OBO relations
- Provide documentation
- Plurality of independent users
- Developed collaboratively

Span of biomedical domain

- Granularity from molecule to organism





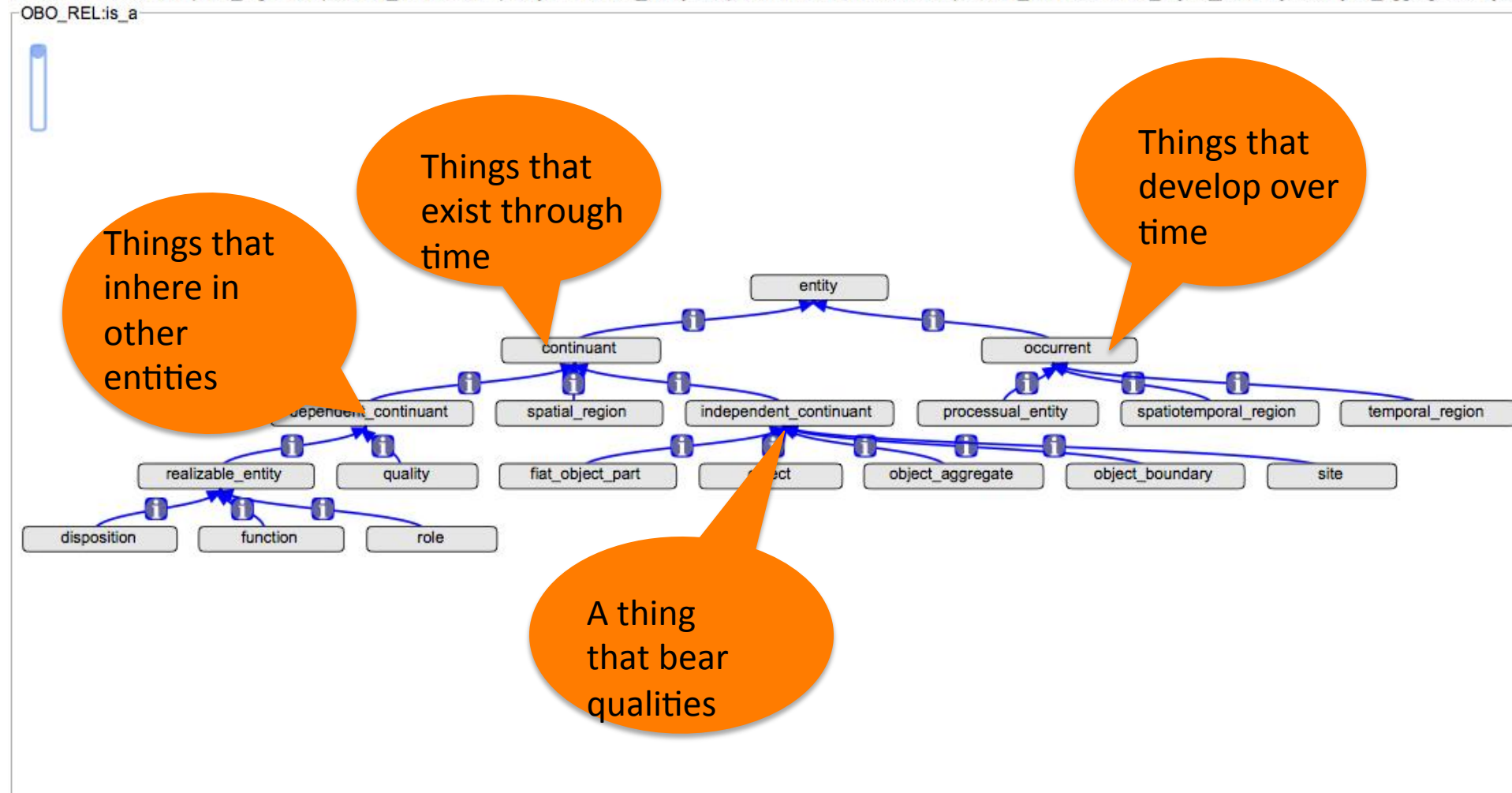
BFO

basic formal ontology

- “narrowly focused on the task of providing a genuine upper ontology which can be used in support of domain ontologies developed for scientific research, as for example in biomedicine within the framework of the [OBO Foundry](http://www.ifomis.org/bfo).”

Top level terms of BFO

Ancestor views of temporal_region, dependent_continuant, quality, realizable_entity, disposition, function, role, independent_continuant, fiat_object_part, object, object_aggregate, obje.





relation ontology

- The RO provides class level and instance level relations.
- ALL SOME RULE – for class level relations
- Improves reasoning
- *All start_codon part_of some CDS*

Questions?

Excellent book

- Introduction to bio-ontologies by Robinson and Bauer
- CRC Press

