High-throughput Sequencing course 2019

Anna Maria Masci

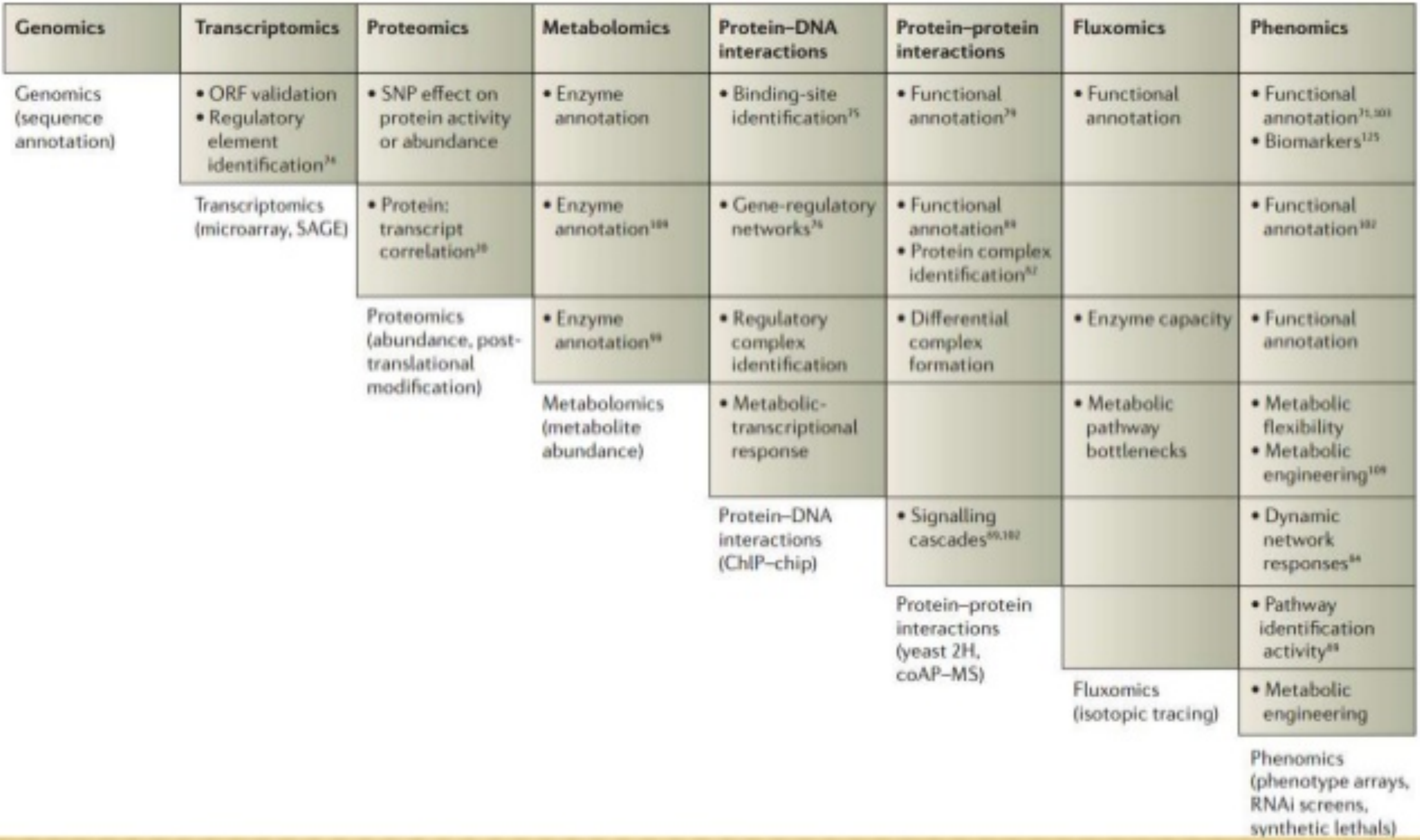
OMICS ERA

What is ‘OMICS’: derived from Latin suffix “ome” meaning mass or many.

Thus OMICS involve a mass (large number) of measurements per endpoint. (Jackson et al. 2006)

- Genomics
- Transcriptomics
- Proteomics
- Reactomics
- Metabonomics
- Phenomics

APPLICATION OF DIFFERENT OMICS



Genomics	Transcriptomics	Proteomics	Metabolomics	Protein–DNA interactions	Protein–protein interactions	Fluxomics	Phenomics
Genomics (sequence annotation)	• ORF validation • Regulatory element identification[36]	• SNP effect on protein activity or abundance	• Enzyme annotation	• Binding-site identification[35]	• Functional annotation[79]	• Functional annotation	• Functional annotation[71,101] • Biomarkers[125]
	Transcriptomics (microarray, SAGE)	• Protein: transcript correlation[38]	• Enzyme annotation[109]	• Gene-regulatory networks[76]	• Functional annotation[88] • Protein complex identification[87]		• Functional annotation[102]
		Proteomics (abundance, post-translational modification)	• Enzyme annotation[99]	• Regulatory complex identification	• Differential complex formation	• Enzyme capacity	• Functional annotation
			Metabolomics (metabolite abundance)	• Metabolic-transcriptional response		• Metabolic pathway bottlenecks	• Metabolic flexibility • Metabolic engineering[109]
				Protein–DNA interactions (ChIP–chip)	• Signalling cascades[69,102]		• Dynamic network responses[84]
					Protein–protein interactions (yeast 2H, coAP–MS)		• Pathway identification activity[88]
						Fluxomics (isotopic tracing)	• Metabolic engineering
							Phenomics (phenotype arrays, RNAi screens, synthetic lethals)

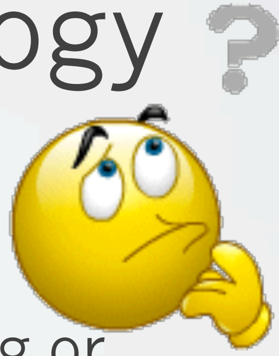
Genomics

Refer to the use of computational and statistical analysis to decipher biology from genome sequences including both DNA and RNA sequences as other post-genomic data (i.e. Genomic microarrays)

Functional Genomics

Is a field of molecular biology that attempts to make use of the vast wealth of data produced by genomics (i.e. genomic sequences) to describe function and interactions. Unlike Genomics, functional genomics focuses on the dynamic aspects such as transcription, translation, protein-protein interactions. Is a more high-throughput methods rather than a gene by gene approach.

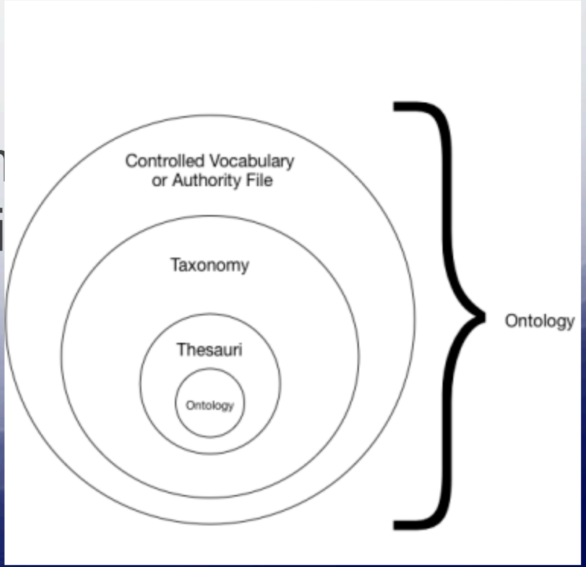
Confusion with the word ontology



Controlled Vocabulary:
Is a restricted list of words used for labeling, indexing or categorizing

Taxonomy:
Is a classification system. The taxonomy in a controlled vocabulary indicates a hierarchical structure based on an IsA relationship

Ontology:
Is a controlled vocabulary with a taxonon
complex than complex than just IsA relati



Using Ontology for data analysis

- Standard identifiers for classes and relations that represent the phenomena within a domain;
- Vocabulary for a domain;
- Metadata that describes the intended meaning of the classes and relations in ontologies;
- Machine-readable axioms and definitions that enable computational access to some aspects of the meaning of classes and relations

The use of standard identifiers for classes and relations in ontologies is a key component in enabling data integration across multiple databases, because the same identifiers can be used across multiple, disconnected databases, files or web sites

These identifiers are widely used in structured file formats, in knowledge bases and data repositories.

Most successful ontology thus far

$200 mill. invested in literature and database curation using the Gene Ontology (GO) since 1999

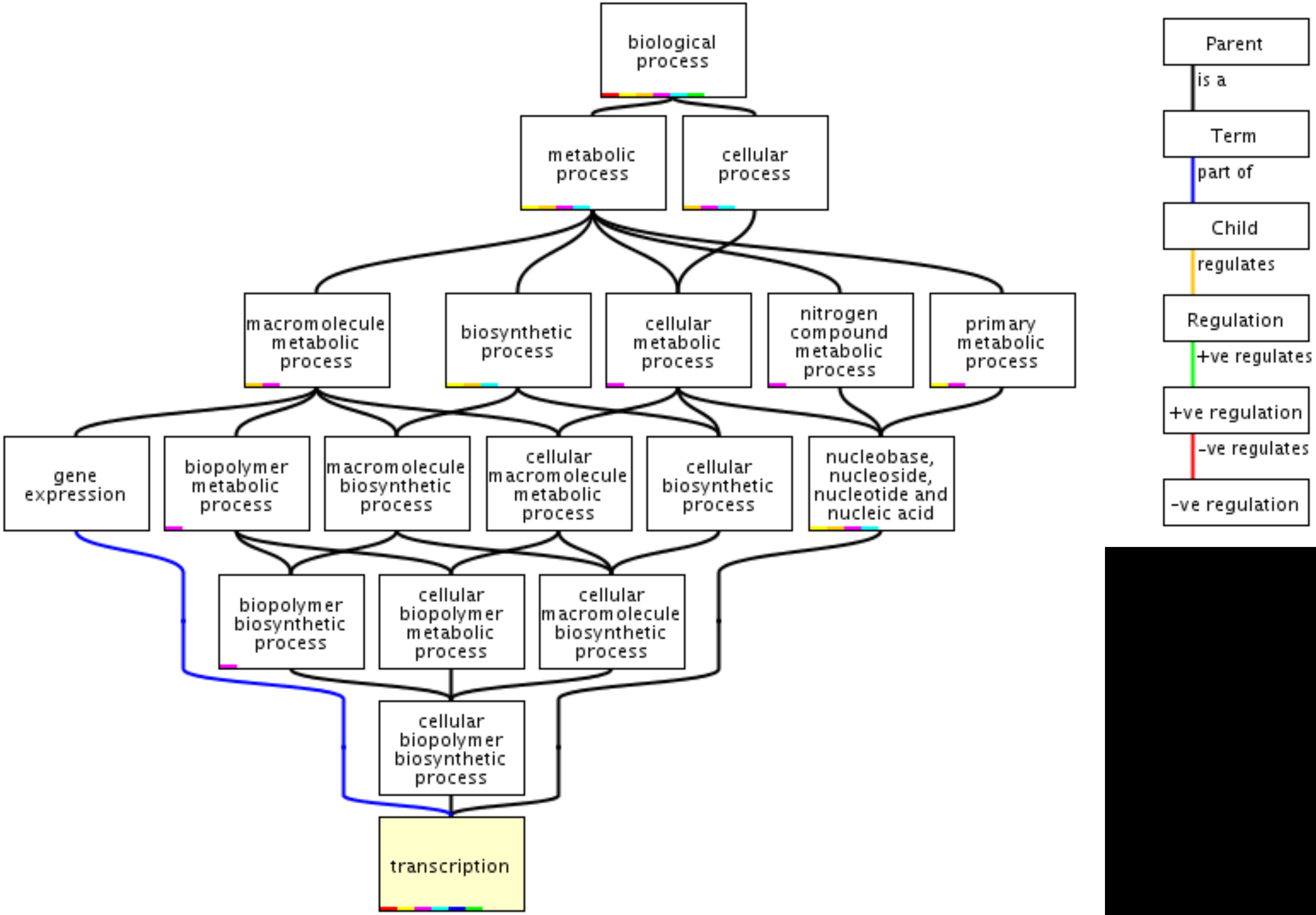
over 11 million annotations relating gene products (proteins) described in the UniProt, Ensembl and other databases to terms in the GO

From Barry Smith: http://ncorwiki.buffalo.edu/index.php/Applied_Ontology

GO provides a controlled system of representations for use in annotating data and literature that is

- multi-species
- multi-disciplinary
- multi-granularity, from molecules to population

From Barry Smith: http://ncorwiki.buffalo.edu/index.php/Applied_Ontology



[redacted]

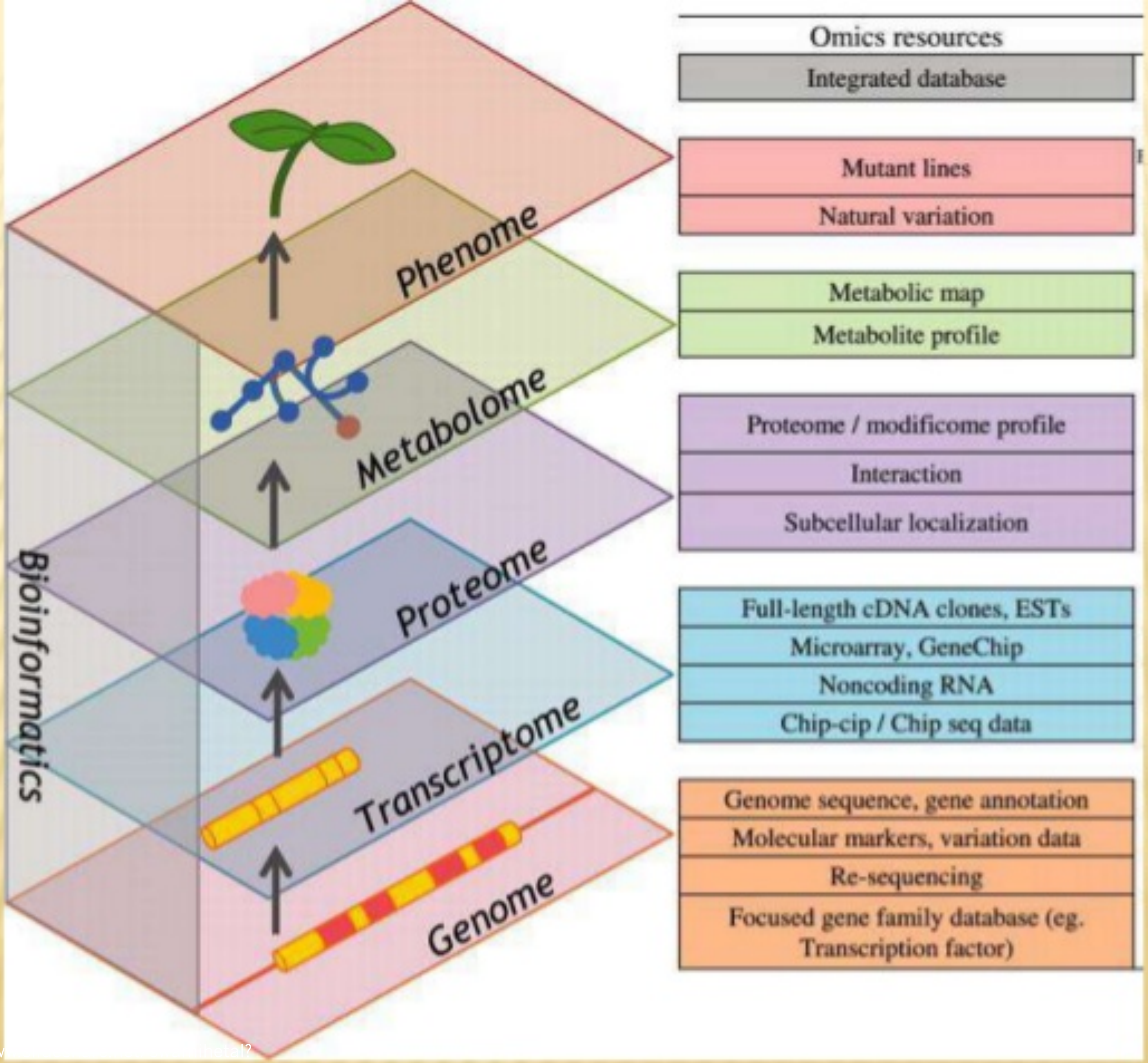
From Barry Smith: http://ncorwiki.buffalo.edu/index.php/Applied_Ontology

Gene ontology structure

GO MF molecular functions
Molecular activities of a gene product

GO BP biological processes
Specific Objective that the organism is genetically programmed to achieve

GO CC cellular components
Where gene products are active



Omics resources
Integrated database
Mutant lines
Natural variation
Metabolic map
Metabolite profile
Proteome / modificome profile
Interaction
Subcellular localization
Full-length cDNA clones, ESTs
Microarray, GeneChip
Noncoding RNA
Chip-cip / Chip seq data
Genome sequence, gene annotation
Molecular markers, variation data
Re-sequencing
Focused gene family database (eg. Transcription factor)

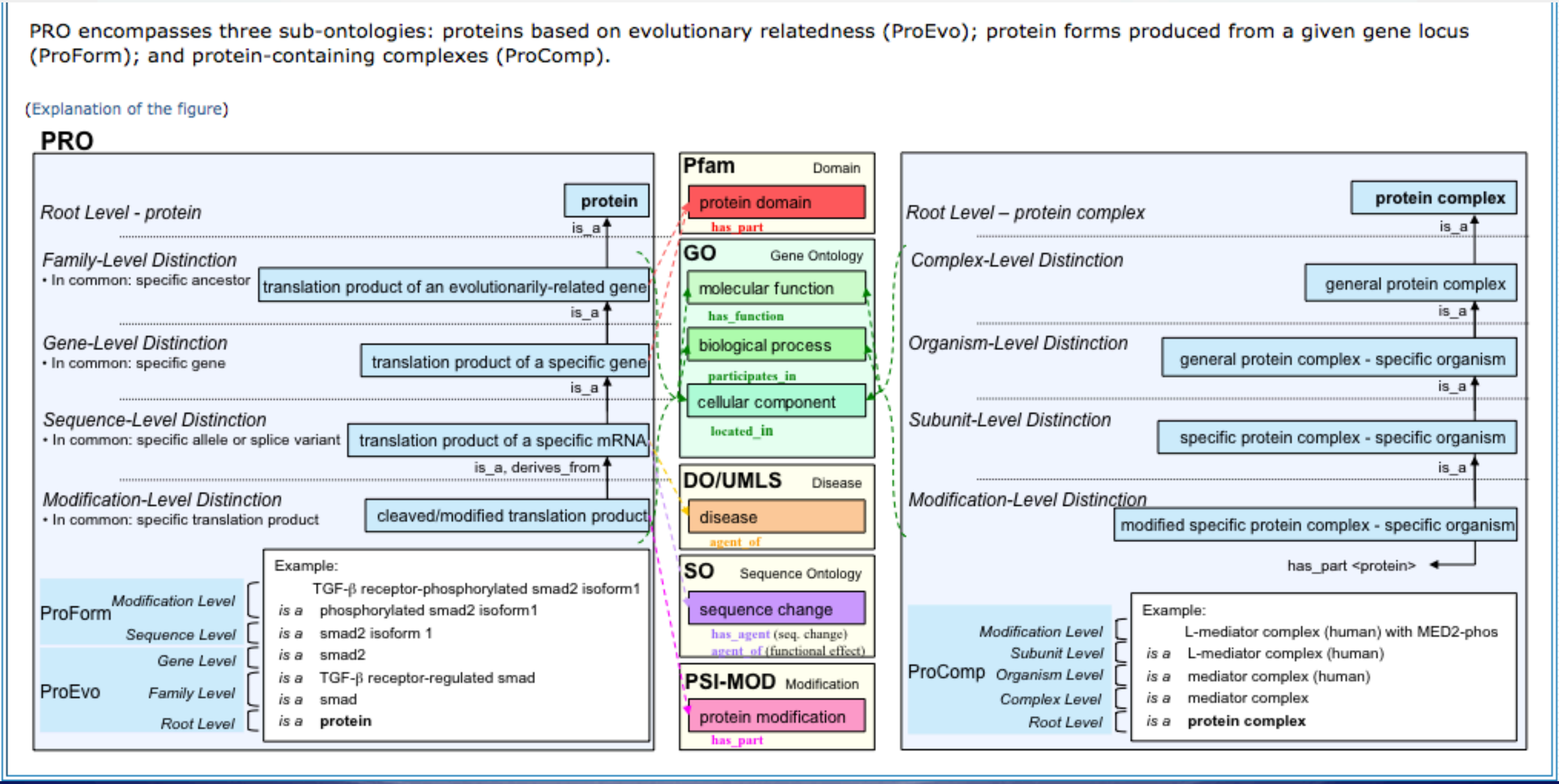
https://ww hetal?
utm_campaign=profiletracking&utm_medium=sssite&utm_source=ssslidevi

High-throughput technologies generated data sets of often thousands of genes, which needed to be interpreted in terms of biological roles.

This provided the impetus behind the ongoing functional and structural annotation of gene products, which is now available through the GO database (AMIGO)and is a mainstay of modern bioinformatics.

Protein Ontology

https://proconsortium.org/



PRO encompasses three sub-ontologies: proteins based on evolutionary relatedness (ProEvo); protein forms produced from a given gene locus (ProForm); and protein-containing complexes (ProComp).

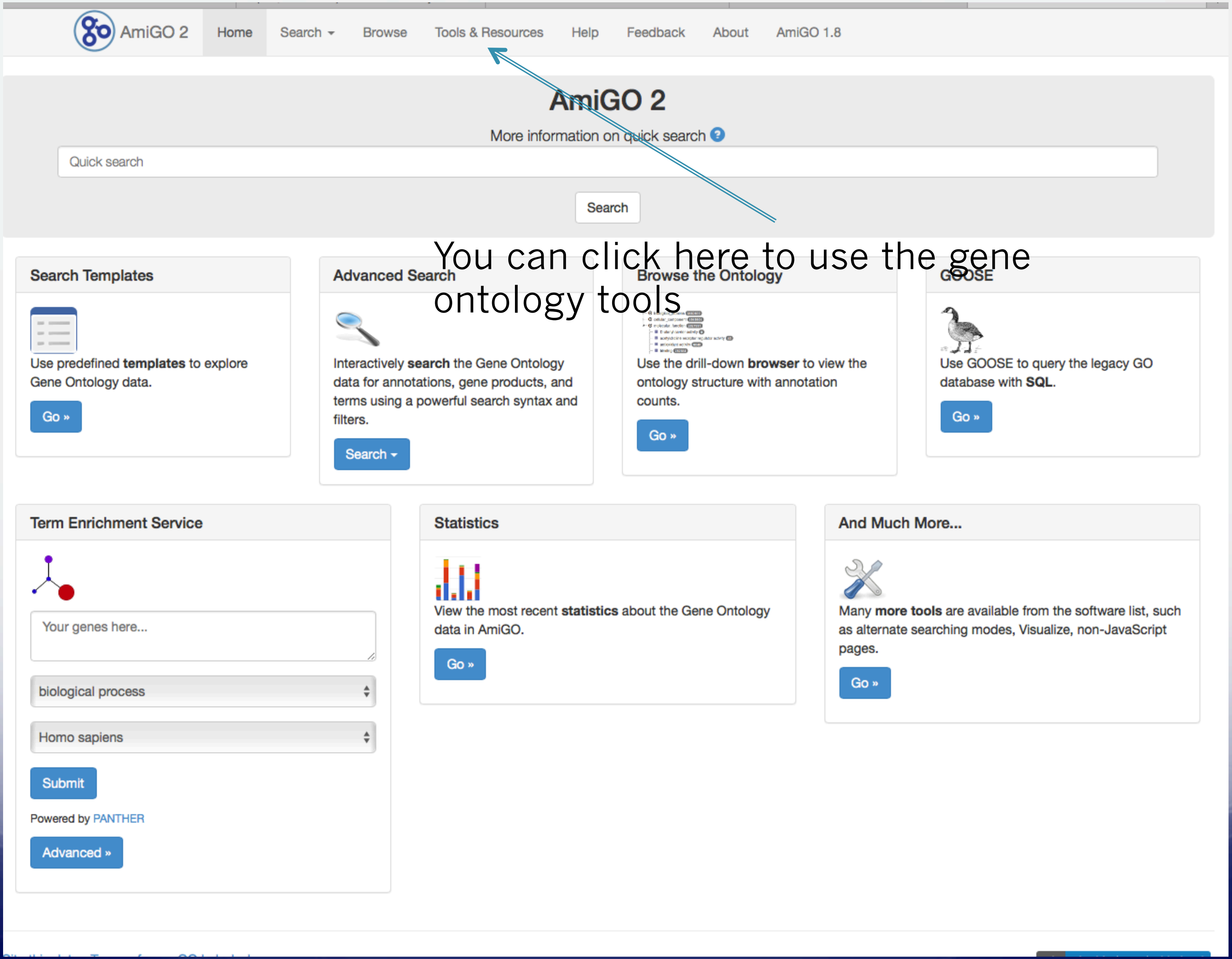
(Explanation of the figure)

Example:
TGF-β receptor-phosphorylated smad2 isoform1
is a phosphorylated smad2 isoform1
is a smad2 isoform 1
is a smad2
is a TGF-β receptor-regulated smad
is a smad
is a **protein**

Example:
L-mediator complex (human) with MED2-phos
is a L-mediator complex (human)
is a mediator complex (human)
is a mediator complex
is a **protein complex**

Go to the AMIGO database and describe it

Amigo2 home page http://amigo.geneontology.org/amigo



You can click here to use the gene ontology tools

AmiGO 2

More information on quick search

Search Templates

Use predefined **templates** to explore Gene Ontology data.

Advanced Search

Interactively **search** the Gene Ontology data for annotations, gene products, and terms using a powerful search syntax and filters.

Browse the Ontology

Use the drill-down **browser** to view the ontology structure with annotation counts.

GOOSE

Use GOOSE to query the legacy GO database with **SQL**.

Term Enrichment Service

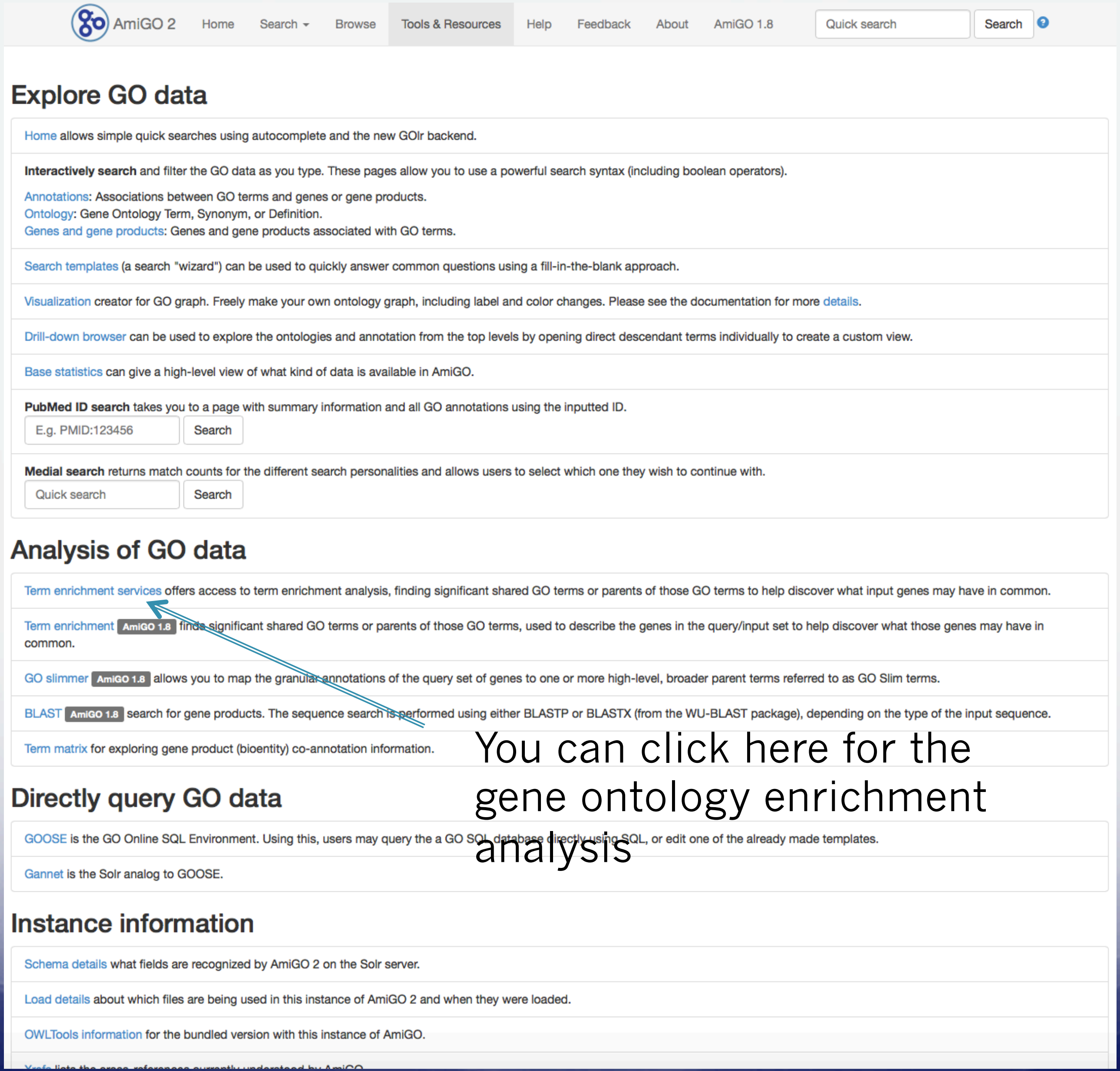
Powered by PANTHER

Statistics

View the most recent **statistics** about the Gene Ontology data in AmiGO.

And Much More...

Many **more tools** are available from the software list, such as alternate searching modes, Visualize, non-JavaScript pages.



Explore GO data

Home allows simple quick searches using autocomplete and the new GOlr backend.

Interactively search and filter the GO data as you type. These pages allow you to use a powerful search syntax (including boolean operators).

Annotations: Associations between GO terms and genes or gene products.
Ontology: Gene Ontology Term, Synonym, or Definition.
Genes and gene products: Genes and gene products associated with GO terms.

Search templates (a search "wizard") can be used to quickly answer common questions using a fill-in-the-blank approach.

Visualization creator for GO graph. Freely make your own ontology graph, including label and color changes. Please see the documentation for more details.

Drill-down browser can be used to explore the ontologies and annotation from the top levels by opening direct descendant terms individually to create a custom view.

Base statistics can give a high-level view of what kind of data is available in AmiGO.

PubMed ID search takes you to a page with summary information and all GO annotations using the inputted ID.

Medial search returns match counts for the different search personalities and allows users to select which one they wish to continue with.

Analysis of GO data

Term enrichment services offers access to term enrichment analysis, finding significant shared GO terms or parents of those GO terms to help discover what input genes may have in common.

Term enrichment AmiGO 1.8 finds significant shared GO terms or parents of those GO terms, used to describe the genes in the query/input set to help discover what those genes may have in common.

GO slimmer AmiGO 1.8 allows you to map the granular annotations of the query set of genes to one or more high-level, broader parent terms referred to as GO Slim terms.

BLAST AmiGO 1.8 search for gene products. The sequence search is performed using either BLASTP or BLASTX (from the WU-BLAST package), depending on the type of the input sequence.

Term matrix for exploring gene product (bioentity) co-annotation information.

You can click here for the gene ontology enrichment analysis

Directly query GO data

GOOSE is the GO Online SQL Environment. Using this, users may query the a GO SQL database directly using SQL, or edit one of the already made templates.

Gannet is the Solr analog to GOOSE.

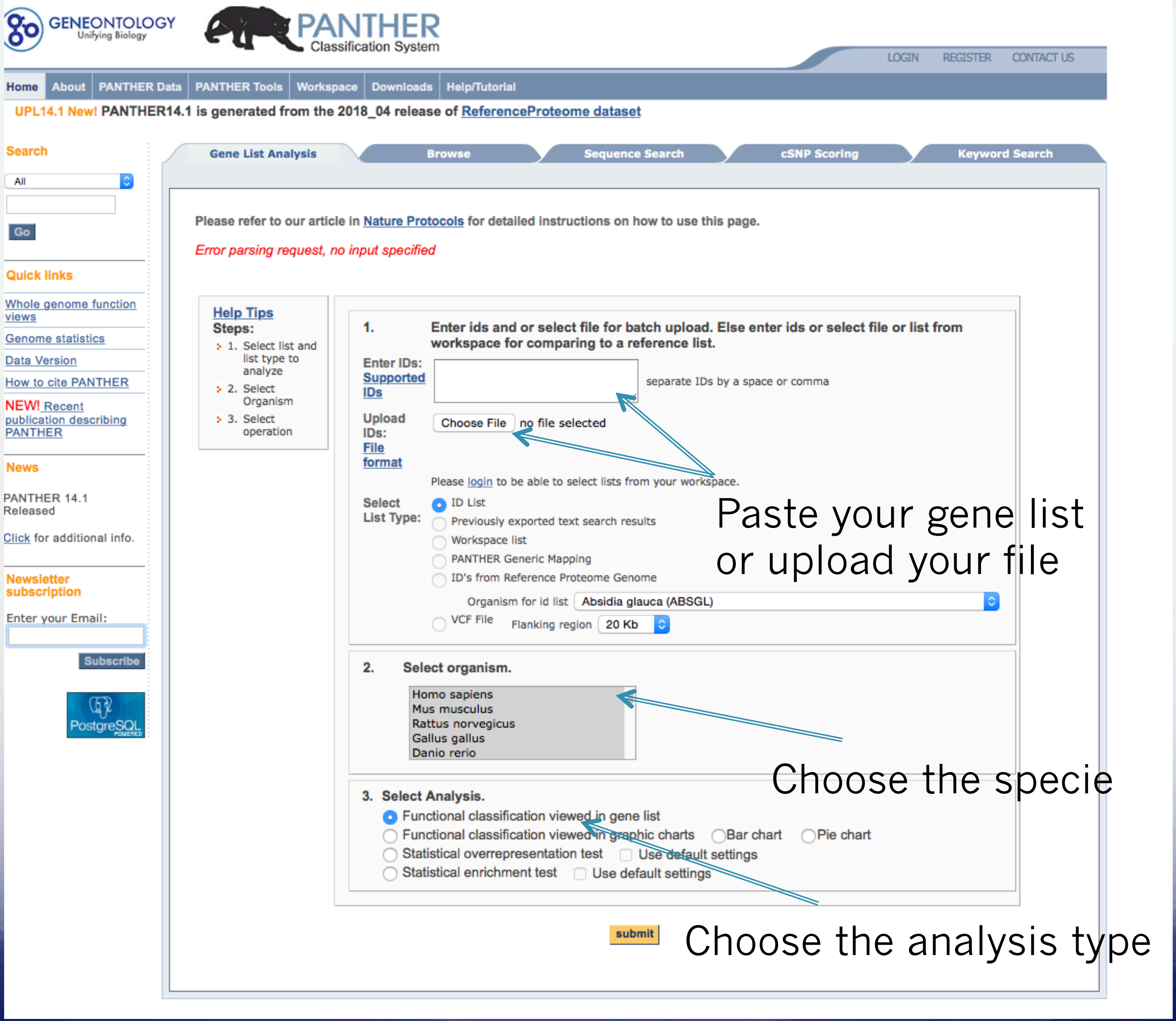
Instance information

Schema details what fields are recognized by AmiGO 2 on the Solr server.

Load details about which files are being used in this instance of AmiGO 2 and when they were loaded.

OWLTools information for the bundled version with this instance of AmiGO.

Xrefs lists the cross-references currently understood by AmiGO.



GENEONTOLOGY Unifying Biology

PANTHER Classification System

LOGIN REGISTER CONTACT US

Home | About | PANTHER Data | PANTHER Tools | Workspace | Downloads | Help/Tutorial

UPL14.1 New! **PANTHER14.1 is generated from the 2018_04 release of ReferenceProteome dataset**

Search

All

Go

Quick links

Whole genome function views

Genome statistics

Data Version

How to cite PANTHER

NEW! Recent publication describing PANTHER

News

PANTHER 14.1 Released

Click for additional info.

Newsletter subscription

Enter your Email:

Subscribe

PostgreSQL

Gene List Analysis | Browse | Sequence Search | cSNP Scoring | Keyword Search

Please refer to our article in Nature Protocols for detailed instructions on how to use this page.

Error parsing request, no input specified

Help Tips

Steps:

1. Select list and list type to analyze
2. Select Organism
3. Select operation

1. **Enter ids and or select file for batch upload. Else enter ids or select file or list from workspace for comparing to a reference list.**

Enter IDs: **Supported IDs** — separate IDs by a space or comma

Upload IDs: **File format** — Choose File no file selected

Please login to be able to select lists from your workspace.

Select List Type:

- ● ID List
- ○ Previously exported text search results
- ○ Workspace list
- ○ PANTHER Generic Mapping
- ○ ID's from Reference Proteome Genome

Organism for id list: Absidia glauca (ABSGL)

○ VCF File — Flanking region: 20 Kb

2. **Select organism.**

Homo sapiens
Mus musculus
Rattus norvegicus
Gallus gallus
Danio rerio

3. **Select Analysis.**

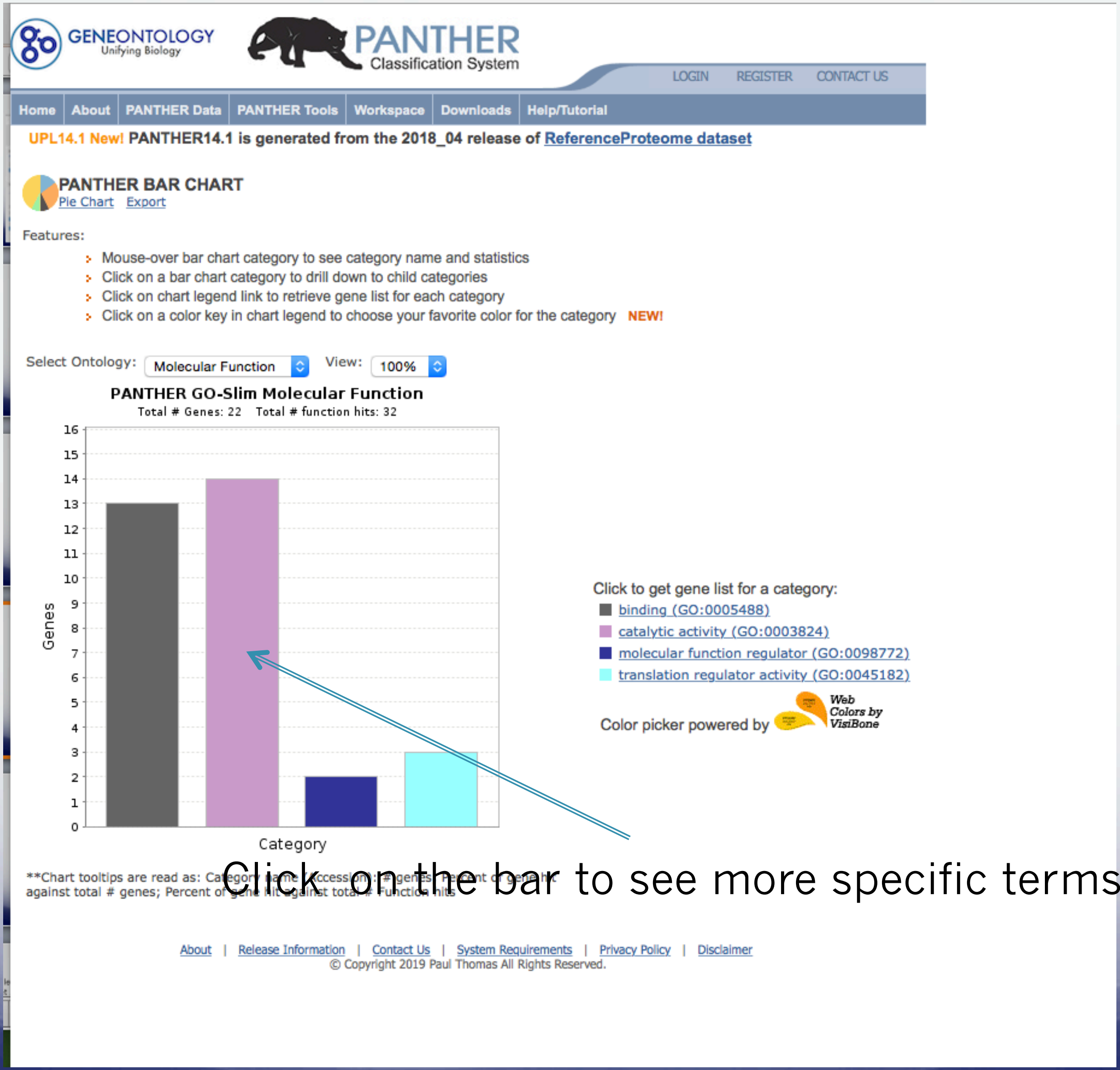
- ● Functional classification viewed in gene list
- ○ Functional classification viewed in graphic charts ○ Bar chart ○ Pie chart
- ○ Statistical overrepresentation test ☐ Use default settings
- ○ Statistical enrichment test ☐ Use default settings

submit

Paste your gene list or upload your file

Choose the specie

Choose the analysis type



UPL14.1 New! PANTHER14.1 is generated from the 2018_04 release of ReferenceProteome dataset

PANTHER BAR CHART

Features:

- Mouse-over bar chart category to see category name and statistics
- Click on a bar chart category to drill down to child categories
- Click on chart legend link to retrieve gene list for each category
- Click on a color key in chart legend to choose your favorite color for the category **NEW!**

PANTHER GO-Slim Molecular Function

Total # Genes: 22 Total # function hits: 32

Click to get gene list for a category:

- ■ binding (GO:0005488)
- ■ catalytic activity (GO:0003824)
- ■ molecular function regulator (GO:0098772)
- ■ translation regulator activity (GO:0045182)

Color picker powered by

**Chart tooltips are read as: Category name (Accession): # genes; Percent of gene hit against total # genes; Percent of gene hit against total # Function hits

About | Release Information | Contact Us | System Requirements | Privacy Policy | Disclaimer

Click on the bar to see more specific terms

Fungi Databases

[Eupath DB](#)

[FungiDB](#)

[YeastMine](#)

Select gene ids and paste your gene list than Get Answer

You will get a list of the Gene Id; organism; transcript IDs; location etc.

You can  Transform to pathway

Find Orthologous

Go to https://yeastmine.yeastgenome.org/yeastmine/begin.do

.

Homolog
A gene related to a second gene by descent from a common ancestral DNA sequence. The term, homolog, may apply to the relationship between genes separated by the event of speciation (see Ortholog) or to the relationship between genes separated by the event of genetic duplication.

Ortholog
Orthologs are genes in different species that evolved from a common ancestral gene by speciation. Normally, orthologs retain the same function in the course of evolution. Identification of orthologs is critical for reliable prediction of gene function in newly sequenced genomes.

Paralog
Paralogs are genes related by duplication within a genome. Orthologs retain the same function in the course of evolution, whereas paralogs evolve new functions, even if these are related to the original one.

STEPS 1

1. Search for genes and organism Cryptococcus neoformans var. grubii H99

2. Paste the list of the all Cryptococcus neoformans var. grubii H99 and **RUN**

3. You will find genes that are present in the database

4. Than do save list a new page will open

STEPS 2

5. Than under Orthologous click S.cerevisiae and you will get a list of orthologous

6. Than under Orthologous you can click on human or mouse orthologous of S.cerevisiae

7. Than you can use that list for search Gene Ontology or other human or mouse databases.

High-throughput technologies generated data sets of often thousands of genes, which needed to be interpreted in terms of biological roles.

This provided the impetus behind the ongoing functional and structural annotation of gene products, which is now available through the GO database (AMIGO)and is a mainstay of modern bioinformatics.

Thank you

Any questions contact me:
annamaria.masci@duke.edu