

## 9 - OBO Language and Gene Ontology

**André Lamúrias**

---

## OBO Language

## Overview

- Open Biomedical Ontologies (OBO) Language
- Developed by the Gene Ontology (GO) Consortium
- Main target the GO
- Adopted by numerous bio-ontologies
- Subset of the semantics of OWL 2 language

## OBO Ontologies

- Composed of
  - A header
  - Provides information about the ontology
  - A set of stanzas
  - Correspond to the content of the ontology

## OBO header

- Information about
  - Format
  - Version date
  - Subsetdef indications - Slims
  - top-level terms
  - Synonyms
  - Name
  - Metaproperties

## OBO header Example

### ■ Gene Ontology

```
format-version: 1.2
data-version: releases/2023-04-01
subsetdef: chebi_ph7_3 "Rhea list of ChEBI terms representing the major species
    at pH 7.3."
subsetdef: gocheck_do_not_annotate "Term not to be used for direct annotation"
...
subsetdef: goslim_yeast "Yeast GO slim"
synonymtypedef: syngo_official_label "label approved by the SynGO project"
synonymtypedef: systematic_synonym "Systematic synonym" EXACT
default-namespace: gene_ontology
ontology: go
property_value: http://purl.org/dc/elements/1.1/description "The Gene Ontology
    (GO) provides a framework and set of concepts for describing the functions
    of gene products from all organisms." xsd:string
property_value: http://purl.org/dc/elements/1.1/title "Gene Ontology" xsd:string
property_value: http://purl.org/dc/terms/license http://creativecommons.org/
    licenses/by/4.0/
property_value: owl:versionInfo "2023-04-01" xsd:string
```

# OBO Language

## OBO stanzas

- Contain key-value lines
- Refer to
  - Universal types - classes/concepts - ([Term])
  - Type definitions - properties/roles - ([TypeDef])
  - Instances - objects/individuals - ([Instance])
- Classified into namespaces
  - For GO:
  - Molecular function
  - Biological process
  - Cellular component

## OBO stanzas

- Alternative identifiers
- Definition of the item
  - with reference to the source(s)/author(s)
- Subsets to which this item belongs
- Different semantic links
  - Synonyms
  - External references
  - Relations between terms
  - Logical definitions
- Comments indicated with !



# OBO Language

## OBO stanzas - Example

```
[Term]
id: GO:0000003
name: reproduction
namespace: biological_process
alt_id: GO:0019952
alt_id: GO:0050876
def: "The production of new individuals that contain some portion of genetic
      material inherited from one or more parent organisms." [GOC:go_curators,
      GOC:isa_complete, GOC:jl, ISBN:0198506732]
subset: goslim_agr
subset: goslim_chembl
subset: goslim_flybase_ribbon
subset: goslim_pir
subset: goslim_plant
synonym: "reproductive physiological process" EXACT []
xref: Wikipedia:Reproduction
is_a: GO:0008150 ! biological_process
disjoint_from: GO:0044848 ! biological phase
```

## OBO stanzas - Example

- Also definitions of relations

```
[Typedef]
id: part_of
name: part of
namespace: external
xref: BFO:0000050
is_transitive: true
inverse_of: has_part ! has part
```

## OBO stanzas - Example

- Such definitions can be used in other stanzas

```
[Term]
id: GO:0000139
name: Golgi membrane
namespace: cellular_component
def: "The lipid bilayer surrounding any of the compartments of the Golgi
      apparatus." [GOC:mah]
is_a: GO:0098588 ! bounding membrane of organelle
relationship: part_of GO:0005794 ! Golgi apparatus
```

## Logical relations

- Subclasses - `is_a`
- Class Disjointness - `disjoint_from`
- Property characteristics
  - Transitivity - `is_transitive`
  - Inverse properties - `inverse_of`
- Logical definitions
  - Set of lines starting with `intersection_of`
  - Equivalent to the conjunction of these terms/representations

## OBO stanzas - Example

- Logical definition

```
[Term]
id: GO:0000019
name: regulation of mitotic recombination
namespace: biological_process
def: "Any process that modulates the frequency, rate or extent of DNA
      recombination during mitosis." [GOC:go_curators]
synonym: "regulation of recombination within rDNA repeats" NARROW []
is_a: GO:0000018 ! regulation of DNA recombination
intersection_of: GO:0065007 ! biological regulation
intersection_of: regulates GO:0006312 ! mitotic recombination
relationship: regulates GO:0006312 ! mitotic recombination
```

## Gene Ontology

## Overview

- Comprehensive model of biological systems
- From the molecular level to larger pathways, cellular and organism-level systems
- Computational representation of scientific knowledge about the function of genes
- Taking into considerations all possible organisms
- Widely used to support scientific research
- Cited in tens of thousands of publications
- Linked to many other biomedical ontologies

# Gene Ontology

## Main Idea

- Understanding gene function is one of the primary aims of biomedical research
- Experimental knowledge obtained in one organism often applicable in others
- If Organisms share relevant genes inherited from common ancestors
- Gene Ontology consortium appeared in 1998 with genom studies of three model organisms
- *Drosophila melanogaster* (fruit fly)
- *Mus musculus* (mouse)
- *Saccharomyces cerevisiae* (baker's yeast)
- Create collaborative classification schema for gene function
- Today extended to thousands of organisms



## Usage Overview

- Cross-species comparisons
- Gene-expression profiling experiments
- Automatic annotation of expression sequence tags (EST) and genomes
- Comparative genomics
- Network modeling
- Analysis of semantic similarity

# Gene Ontology

## Three Subontologies

### ■ Molecular Function

- Biochemical activity of a gene product
- On a molecular level of granularity
- No indication when or where event occurs (or purpose)

Term	Term ID	Definition
mannosyltransferase activity	GO:0000030	Catalysis of the transfer of a mannosyl group to an acceptor molecule, typically another carbohydrate or a lipid
zinc binding	GO:0008270	Interacting selectively and noncovalently with zinc (Zn) ions

# Gene Ontology

## Three Subontologies

### ■ Biological Process

- Biological objective to which gene (product) contributes
- Assemblies of molecular function, collection of events with beginning and end
- At the level of granularity of the cell or organism

Term	Term ID	Definition
ossification	GO:0001503	The formation of bone or of a bony substance, or the conversion of fibrous tissue or of cartilage into bone or a bony substance
regulation of glial cell proliferation	GO:0060251	Any process that modulates the frequency, rate or extent of glial cell proliferation

# Gene Ontology

## Three Subontologies

- Cellular Component
  - Location where gene product is active

Term	Term ID	Definition
Golgi apparatus	GO:0005794	A compound membranous cytoplasmic organelle of eukaryotic cells, consisting of flattened, ribosome-free vesicles arranged in a more or less regular stack
viral capsid	GO:0019208	The protein coat that surrounds the infective nucleic acid in some virus particles

## Relations

- Subclasses *is\_a*
- Part-whole relations *part\_of*
  - E.g. nucleus is part of a cell  $nucleus \sqsubseteq \exists part\_of. cell$
- Relations between processes *regulates*
  - With subrelations *positively\_regulates* and *negatively\_regulates*
- Whole-part relations *has\_part*
  - Inverse to *part\_of*
  - But the established relations are not necessary inverse
$$nucleus \sqsubseteq \exists has\_part. chromosome$$
  - Does not imply that every chromosome is part of a nucleus

## Annotations

# Annotations

## Overview

- Used during curation process
- Terms of GO do not refer to specific genes
- Rather to their characteristics
- Annotation indicates that a GO term applies to a particular gene product
- Biocurators read full-text articles in their area of expertise and add information to a database using structured vocabularies, such as GO

# Annotations

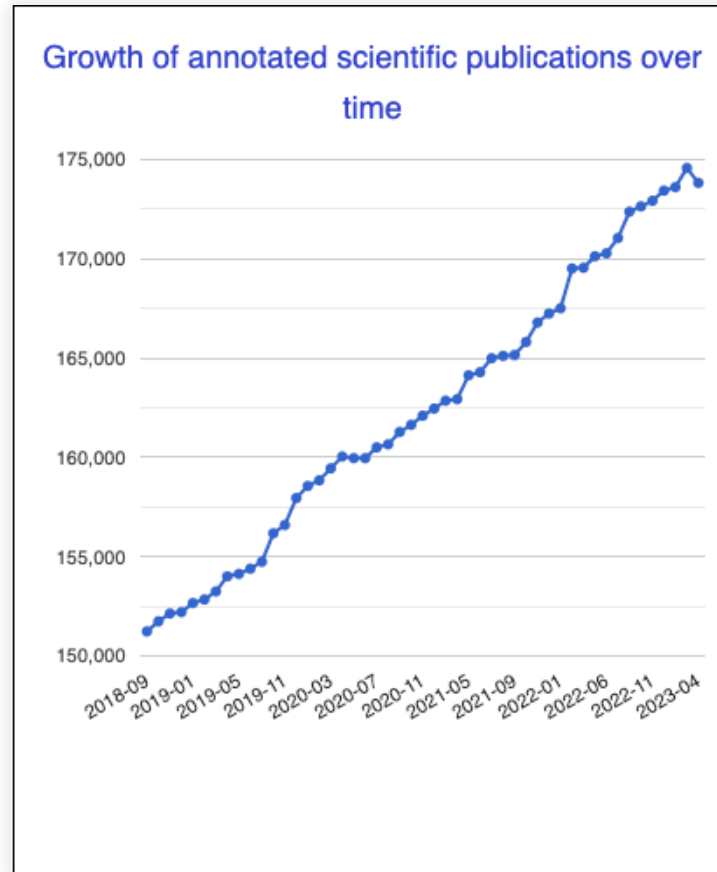
## Some Statistics for GO

- Release April 2023
- Number of annotations: 7,442,411
- Number of annotated scientific publications: 173,800
- Annotated gene products: 1,502,221
- Annotated species: 5291



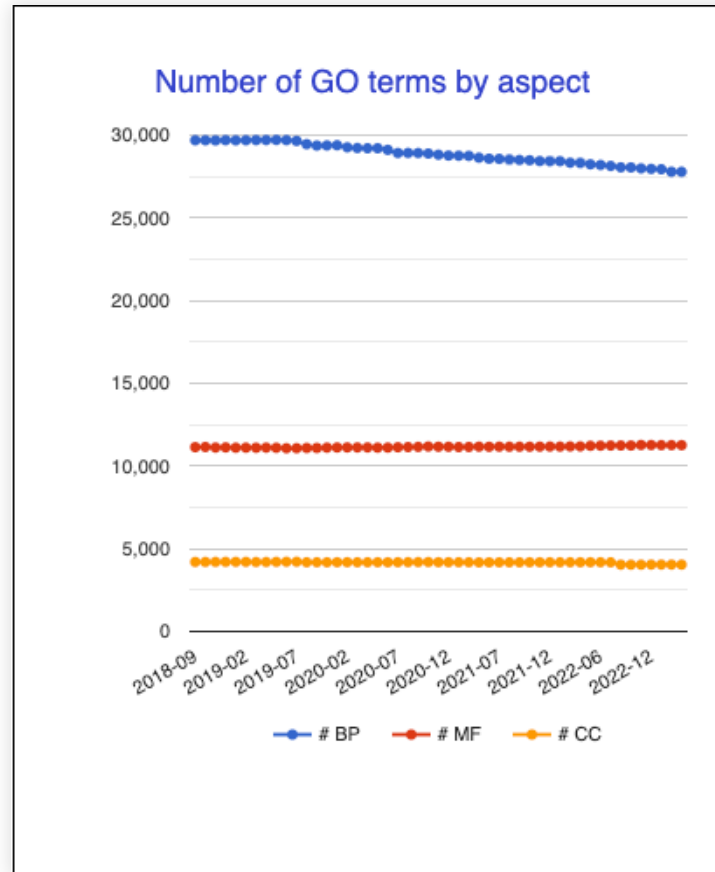
# Annotations

## Growth of Annotated Publications



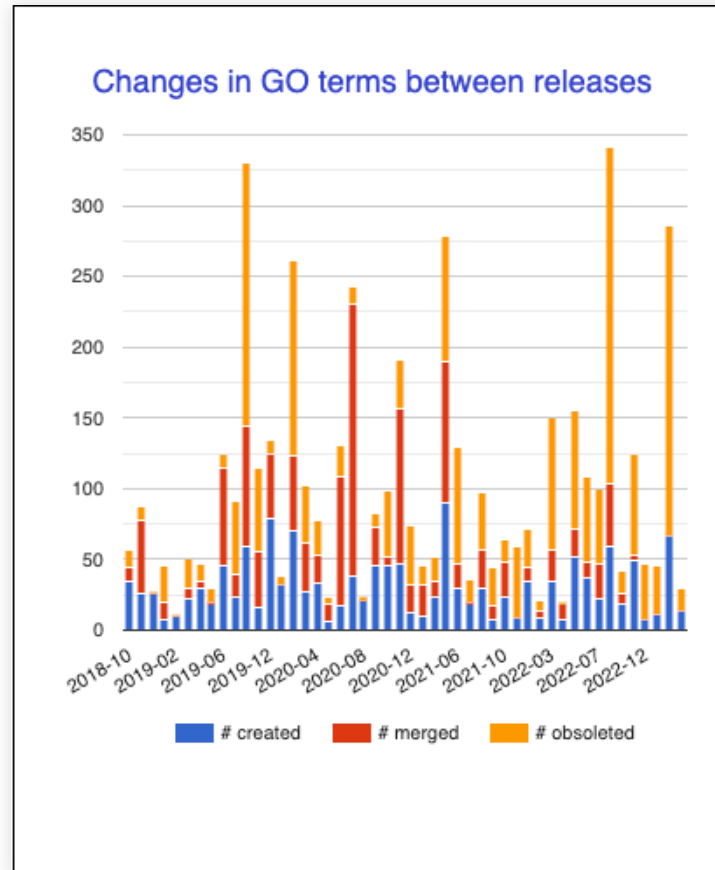
# Annotations

## Terms per Subontology



# Annotations

## Changes in GO terms



# Annotations

## Annotation information

- Gene identifier
- GO term
- Type of evidence to support annotation
- reference to the evidence
- Further complementary data (on database, synonyms, species etc.)

## Evidence for Gene Functions

- Tradeoff between coverage and accuracy
- Highest-quality annotations from experiments
- Computational analysis - based on in silico analysis
- Sequence orthology - genes in two different species have common evolutionary origin
- E.g., because of that location/type of gene in a species can be inferred
- Author statement evidence
- For references to articles that refer to papers with the original research
- Curatorial evidence based on inference by curator from GO annotations
- ND - no biological data available

## Inferred from Electronic Annotations

- Generated automatically, and not validated (yet)
- Two kinds:
  - Map functional data from other databases with different but compatible vocabulary
  - E.g. UniProt in molecular biology with different goals (compared to GO)
  - MGI map UniProt keywords to GO terms to create GO annotations automatically
  - Use common origin of genes to pass annotations from well-studied organisms to less-studied organisms

# Annotations

## General Principles (from GO webpage)

- Annotations represent normal functions of gene products
- A gene product annotated to zero or more terms from each ontology
- Each annotation supported by a GO evidence code from the evidence and conclusions ontology and a references
- Gene products annotated to the most granular term in the ontology that is supported by available evidence
- Annotation to a GO term implies annotation to all its parents
- Annotations may change over time and reflect the current view
- Open world assumption - lack of annotation - role still unknown

# OBO Language and Gene Ontology

## Summary

- OBO Language
- Gene Ontology
- Annotations

## Further reading:

- Robinson and Bauer, Introduction to Bio-Ontologies, Chapters 4.1 and 5
- Dessimoz and Skunca, The Gene Ontology Handbook
- GO webpage <http://geneontology.org/>



