Connecting Function and Process

It’s simpler than you think
Originally GO consisted of 3 ontologies that describe the what and where of a gene products actions

Molecular Function
GO term: Malate dehydrogenase.
GO id: GO:0030060
(S)-malate + NAD(+) = oxaloacetate + NADH.

Biological Process
GO term: developmental process
Synonym: development
GO id: GO:0032502
Definition: A biological process whose specific outcome is the progression of an integrated living unit: an anatomical structure (which may be a subcellular structure, cell, tissue, or organ), or organism over time from an initial condition to a later condition.

Cellular Component
GO term: mitochondrion
GO id: GO:0005739
Definition: A semiautonomous, self replicating organelle that occurs in varying numbers, shapes, and sizes in the cytoplasm of virtually all eukaryotic cells. It is notably the site of tissue respiration.
Three ontologies designed for practical purposes

• Cellular component
  – The part of a cell or its extracellular environment in which a gene product is located. A gene product may be located in one or more parts of a cell and its location may be as specific as a particular macromolecular complex, that is, a stable, persistent association of macromolecules that function together.

• Biological Process
  – Any process specifically pertinent to the functioning of integrated living units: cells, tissues, organs, and organisms. A process is a collection of molecular events with a defined beginning and end.

• Molecular Function
  – Elemental activities, such as catalysis or binding, describing the actions of a gene product at the molecular level. A given gene product may exhibit one or more molecular functions.
From the original ‘Nature Genetics’ Paper

‘Biological process refers to a biological objective to which the gene or gene product contributes. A process is accomplished via one or more ordered assemblies of molecular functions.’
Why three ontologies?

• Original GO had only part_of and is_a relationships.
• The same molecular functions can be used in different biological processes
  – Using is_a and part_of it was easier to keep them separate
• Molecular functions and some biological processes happen in cellular components
But we always knew that there were relationships between all of these

*hydrogen ion transporting ATP synthase activity, rotational mechanism*

*Suzi: Some problems come back to the fact that there is a relationship between function and process which we don't reflect."

September 2003, Bar Harbor

- Gene product
- Molecular function
- Cellular component
- Biological process

ATP biosynthesis

**ATP5O**
ATP5O

Gene product
Molecular function
Cellular component
Biological process

hydrogen ion transporting ATP synthase activity, rotational mechanism

mitochondrion

ATP biosynthesis

Ontology
We already had relationships within each ontology

• Part_of
  – When it exists it is always a part_of its parent
  – Along with other part_of (s) it makes the parent
• Is_a
  – Has all of the attributes of the parent, but differs in some way from other is_a (s)
• Regulates (added recently)
  – Changes the rate, frequency or extent of something happening
• Has_part (added recently)
  – When it exists is always has the child as a part
GO developers use these relationships all the time
Relationships define the behavior of annotations

The rhomboid gene product is involved in some kind of epidermal growth factor ligand processing.

PMID:11672524
Therefore...

The rhomboid gene product is involved in part of some kind of epidermal growth factor receptor signaling pathway.

PMID: 11672524
Therefore...

The rhomboid gene product is involved in part of some kind of transmembrane receptor protein tyrosine kinase signaling pathway.

PMID:11672524
rhomboid gene product

PMID:11672524

IGI

Etc, Etc, Etc...

biological_process

signaling

signaling pathway

intracellular signaling pathway

enzyme linked receptor protein signaling pathway

transmembrane receptor protein tyrosine kinase signaling pathway

epidermal growth factor receptor signaling pathway

epidermal growth factor ligand processing
So, back to functions and processes

hydrogen ion transporting ATP synthase activity, rotational mechanism

ATP5O

ATP biosynthesis

mitochondrion

Gene product
Molecular function
Cellular component
Biological process
So, back to functions and processes

ATP5O

hydrogen ion transporting ATP synthase activity, rotational mechanism

mitochondrion

Gene product
Molecular function
Cellular component
Biological process

part_of

ATP biosynthesis
We have decided that we are going to make the relationships between ontologies in GO.

Molecular functions are part of biological processes.
When a function happens, it is part of some process

• Keep in mind that to have a part_of relationship to a process in GO, it always has to be a part of that process

• Remember that functions can be, and often are, used over and over again. That is one of the reasons functions and processes were split out in the beginning.

• This was first formally introduced at the Princeton GO meeting, 2007.
This is not a mapping between resources or anything fancy, it is normal ontology development.
So annotations behave the same way!

The *Egfr* gene product is involved in part of some kind of Epidermal growth factor receptor signaling pathway.
What about functions that are used over and over?

Over the years we have run into the same thing within biological process

Within biological process, we have conventionally created subtypes of processes that are used as parts of other processes.

Wnt receptor signaling pathway involved in somitogenesis
We create specific functions the same way we create specific processes.

The first process-specific function
Jesus college, 2009
However, databases can’t handle the function-process links

So Chris has created a file that contains the biological process annotations that are inherited from molecular function annotations.
So how should we change our annotation habits?

- We shouldn’t
- If we all incorporate the new annotation sets, then there is no need to make an annotation to a biological process based on an experiment that is part_of that process.
- If we don’t incorporate the new annotation set then we should make the ‘redundant’ annotation
- Once databases have the ability to use the inter-ontology links, these annotations will propagate like any other annotations.