Annotating to GO regulation terms

Regulation definition

- Any process that modulates the rate, frequency or extent of X
  - X is (generally) something that happens: a process or a function
  - If Y regulates X, it starts it, stops it, slows it down or speeds it up
- The key is *defining* X
  - If we can define X with respect to where it starts, stops and its parts, then it should be easy to tell if Y regulates X

- *If Y regulates X, then it stops it, starts it, slows it down or speeds it up*
- How does process Y stop, start, speed up or slow down process X?
  - It has an effect on the beginning, end or some part of the process
- What makes up the beginning, parts or end of a biological process?

Function-process links

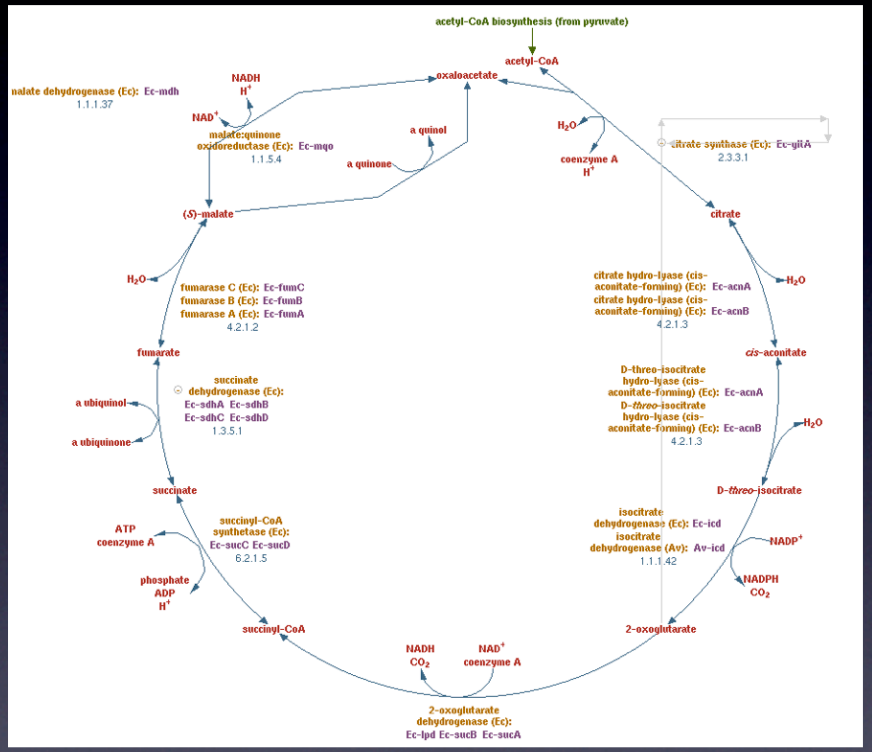
- Biological processes are ordered assemblies of molecular functions
- One of the ongoing tasks of the GO editors is to make part_of links between biological processes and their constituent molecular functions

figure of mf-bp links

- Under this model, one function starts a process, one stops it and their is a series in between
- If another process modulates any one of those functions, then it *regulates* the process
- BUT, the definition of the process is subjective
  - GO needs to reflect the community consensus about which functions are part of a process and which are not

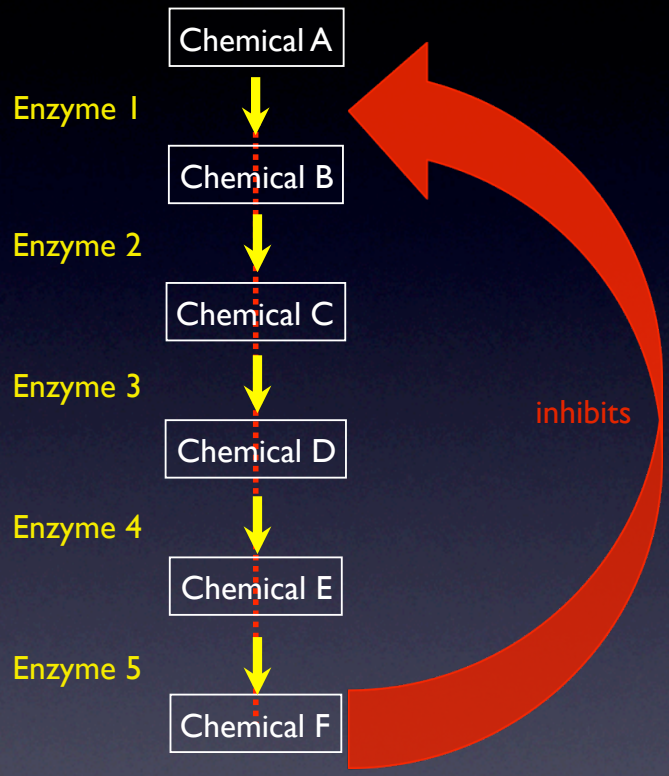
TCA cycle

Sometimes it will be easy



Sometimes it will be impossible

In this case we could say that enzyme 5 is both *part_of* the pathway, and *negatively_regulates* the pathway



- Even for the easy ones, it is a huge amount of work for the ontology developers to define biological processes based on their molecular functions
  - Different pathways in different organisms
  - Need for process-specific functions (why process was separate from function in the first place)
- But this is worth doing because of the inferences it allows us to make, and the clarity it provides

Guidelines for annotation

1. A process is composed of a series of molecular functions:
   - If you know that your gene product performs one of the functions in the process, then the annotation should be to the process term.
   - If you know that your gene product regulates one of the functions, the annotation should be to the regulation term.
   - If you don't know exactly how your gene product is involved in the process, annotate to the process term, which is broader.

2. Use your *biological knowledge*. If it is a well-known pathway and hasn't been fully represented in GO, then background knowledge is needed to decide if the function is *part_of* the pathway, *regulates* the pathway, or does both. Things to consider:
    - How much is known about the process?
    - Is there a defined pathway for this process in which the major players have been identified?
    - Is the gene product being annotated believed to be a major player in the process or pathway or outside of it?

3. Gene products that are a constituent *part_of* a process should only be annotated to regulation of that process where they regulate a *different function* in that process (e.g. by negative feedback), but *not* if it's just their presence that is limiting (e.g. levels of a receptor on a cell surface).

3. If the gene product or pathway is not fully described, then try to *reflect what the paper you are reading is saying*. The author should give hints about what is happening, and will be the experts in this field.

4. Processes should have a *defined beginning and end*. If this isn't clear from the definition, then you may need to start up a dialogue (tracker, mailing lists) to get the term *redefined*.

5. If new information changes the view of a gene product's role in a process, older annotations should be checked and possibly removed for consistency. *Annotations should reflect the most up-to-date view of a gene product's role.*

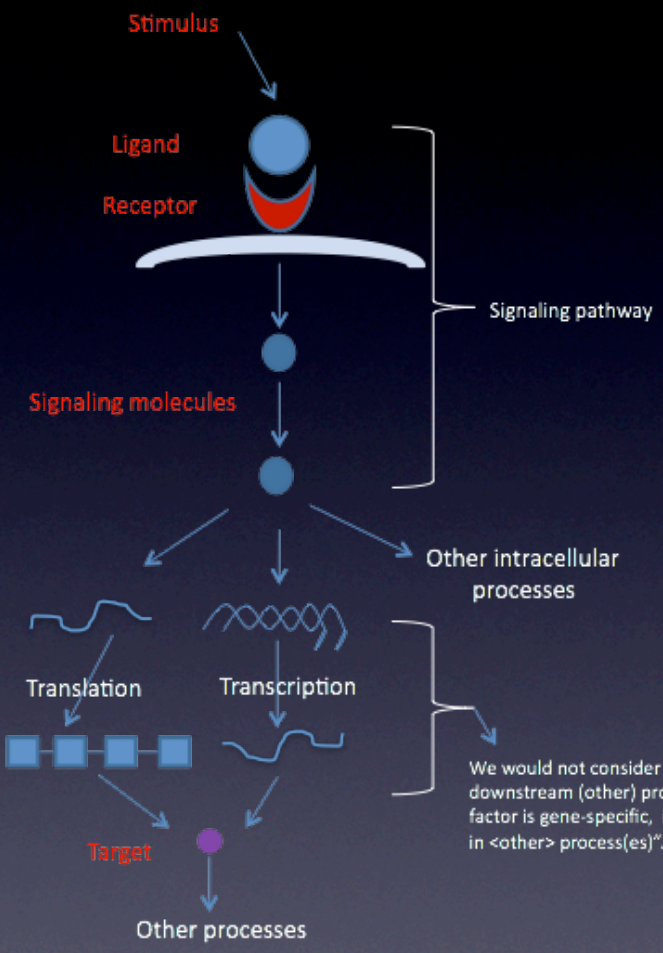
6. Inferred from Mutant Phenotype (IMP)

- When deciding whether to annotate to a parent process term versus a regulation term based upon a mutant phenotype, curators should consider:
  - the assay
  - the nature of the allele used in the studies (null versus reduction-of-function)
  - perhaps also the identity of the gene product to choose an appropriate annotation.
- Unless it is clear that a specific function in the pathway is being regulated in the pathway, the annotation should be made to the pathway term.
  - It's very difficult to make an annotation to a regulation term based solely on a mutant phenotype, so be very careful when making this type of annotation (see example).

Issues

- These guidelines are high-level, annotators need guidance for making calls on individual experiments
  - Use example annotations?
- The beginning and end of processes are not clear
  - For example, the signaling group has decided that in general ligand binding to a receptor is *part_of* that pathway, however, a ligand is likely to be the rate limiting step in a signaling pathway and therefore will be annotated to both the signaling pathway and the regulation of the signaling pathway
    - some gene products involved in synthesis, transport, etc. of ligand which *trigger* pathway X could be annotated to regulation of pathway x
    - members of the pathway would be annotated directly to pathway x
    - downstream effects would be regulated by the pathway, but not *part_of* the pathway

General ligand-receptor pathway



Suggested biological process annotations

Stimulus (intracellular pathway):
Regulation of signaling pathway

Ligand (intracellular pathway):
Signaling pathway
Regulation of <other> process(es)

Ligand (intercellular pathway):
Regulation of signaling pathway
Regulation of <other> process(es)

Receptor:
Signaling pathway
Regulation of <other> process(es)
Cellular response to stimulus/ligand

Signaling molecules:
Signaling pathway
Regulation of gene-specific transcription
Regulation of translation
(Regulation) of transcription in response to <stimulus/ligand>
(Regulation) of transcription involved in <other> process(es)
(Regulation) of <other intracellular> process(es)
Cellular response to stimulus/ligand

Target:
Cellular response to stimulus
<Other> process(es)
Regulation of <other> process(es)

We would not consider annotating the core transcription machinery to the downstream (other) processes that the target is involved in unless the transcription factor is gene-specific, in which case we would annotate to "regulation of transcription involved in <other> process(es)".

Issues

- Re-annotation required?
  - At least 2500 cases where a gp is annotated to both a process *and* its regulation
  - Some of these may be correct (as in negative feedback example) but many will need revisiting
  - How do we handle this? Send lists to individual MODs to check?