Regulation

What we know and what we don’t, annotation same as it ever was

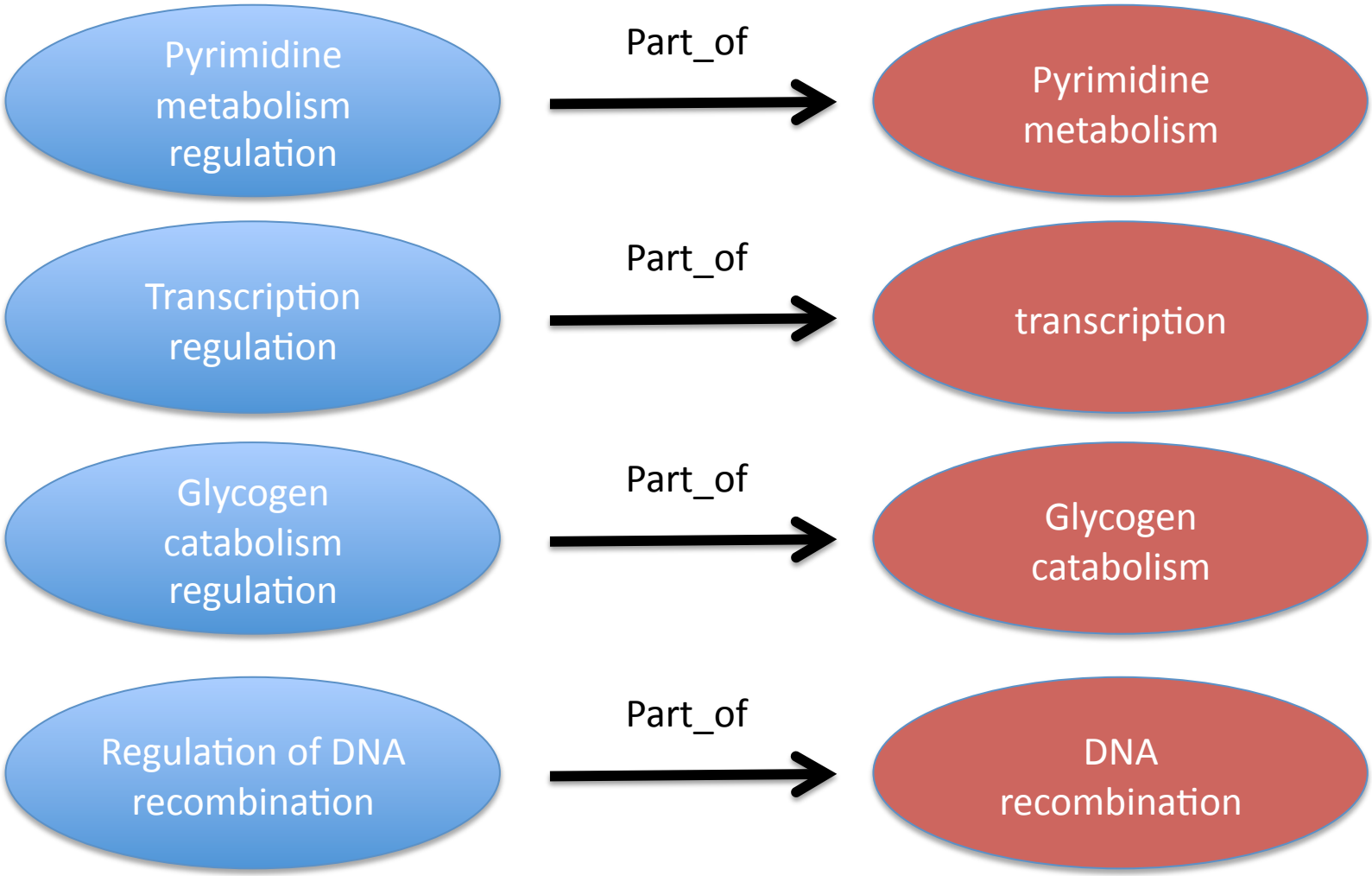
Regulation terms have always existed in GO

From January 1st, 2001 process ontology

<glycogen biosynthesis regulation ; GO:0005979 % carbohydrate biosynthesis ; GO:0016051
<glycogen catabolism regulation ; GO:0005981 % carbohydrate catabolism ; GO:0016052
%general regulation of carbohydrate metabolism ; GO:0006109
%glycolysis regulation ; GO:0006110
%gluconeogenesis regulation ; GO:0006111
%purine metabolism regulation ; GO:0006141
%pyrimidine metabolism regulation ; GO:0006142
%nucleoside metabolism regulation ; GO:0009118
%nucleotide metabolism regulation ; GO:0006140
<DNA replication regulation ; GO:0006275
<DNA repair regulation ; GO:0006282
<regulation of DNA recombination ; GO:0000018
%regulation of mitotic recombination ; GO:0000019 < mitotic recombination ; GO:0006312
%S-phase regulated histone modification ; GO:0006324
%positive regulation of homeotic gene (trithorax group) ; GO:0006339
%negative regulation of homeotic gene (Polycomb group) ; GO:0006340
<transcription regulation ; GO:0006355
%pheromone regulation of gene expression ; GO:0009373 % pheromone response ; GO:0007328
%transcription regulation from Pol I promoter ; GO:0006356
%transcription regulation from Pol II promoter ; GO:0006357
%global transcription regulation from Pol II promoter ; GO:0006358
%transcription regulation from Pol III promoter ; GO:0006359
<general regulation of protein biosynthesis ; GO:0006417
<translational regulation ; GO:0006445
%translational regulation, initiation ; GO:0006446
%iron regulation ; GO:0006447<glycogen biosynthesis regulation ; GO:0005979 % carbohydrate biosynthesis ; GO:0016051
<glycogen catabolism regulation ; GO:0005981 % carbohydrate catabolism ; GO:0016052
%general regulation of carbohydrate metabolism ; GO:0006109

Initially, these terms were a part_of the thing that was being regulated

usually



And so did many others

We knew this wasn't quite right, but at the time we only had 2 relationships

Annotations were made consistently to the most granular term in the ontology that could be assigned based on the literature.

September 2003, Bar Harbor

"Problem 1. There is an ontology problem in that when the function ontology has an enzyme activity and with children "regulation of activity" and "catalytic activity" there becomes a true path violation for the regulator in that it's path goes up to the catalytic activity when it does not have that activity."

If a paper showed that a gene product served a regulatory role in a process

Fibroblast Growth Factor 21 Controls Glycemia via Regulation of Hepatic Glucose Flux and Insulin Sensitivity

Eric D. Berglund, Candice Y. Li, Holly A. Bina, Sara E. Lynes, M. Dodson Michael, Armen B. Shanafelt, Alexei Kharitonenkov, and David H. Wasserman

Then we annotated to the regulation term

If a paper showed that a gene product played a role in part of a process

Developmental Biology 338 (2010) 183–192

Contents lists available at ScienceDirect

Developmental Biology

journal homepage: www.elsevier.com/developmentalbiology

Fgf8b-containing spliceforms, but not *Fgf8a*, are essential for *Fgf8* function during development of the midbrain and cerebellum

Qiuxia Guo, Kairong Li, N. Abimbola Sunmonu, James Y.H. Li *

Department of Genetics and Developmental Biology, University of Connecticut Health Center, Farmington, CT 06030, USA

Then we annotated to the core process term

In summer 2007, we decided to address regulation not always being a part_of a process

After a lot of work parsing terms, fixing term names and fixing relationships, the ‘regulates’ relationships were introduced to GO.

This work was presented at the 2008 Salt Lake City GOC meeting.

We have also introduced a has_part relationship

There should be no effect on annotation practices

In essence, there is no need to change annotation practices, annotators should still annotate to the most granular term possible

This means annotators need to read the paper, understand the biology as best they can, and DECIDE which term is the best.

Sometimes this isn't straightforward

- That's why they pay us the big bucks.
- What if it isn't really clear from a phenotype if a gene product is carrying out a function that is regulating a process or is carrying out a function that is an integral part of the process?
- The fundamental issue here is that we don't understand all of biology

Over the last three years, we have thought A LOT about what regulation means in biology

- The regulates relationship is hard to define in biology.
  - Human nature: people like to show that what they are working on is important
  - Biology is essentially a series of regulatory events, when it comes right down to it, almost everything that happens is regulatory
- But that's not good enough for the ontology
  - Can we define when it's regulatory and when it's a part_of?

Some guidelines about using the regulation terms in GO

- Any process that modulates the rate, frequency or extent of 'X'.
  - 'X' is something that happens, a process or a function
  - If 'Y' regulates 'X', then it stops it, starts it, slows it down, or speeds it up.
- The key here 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'.

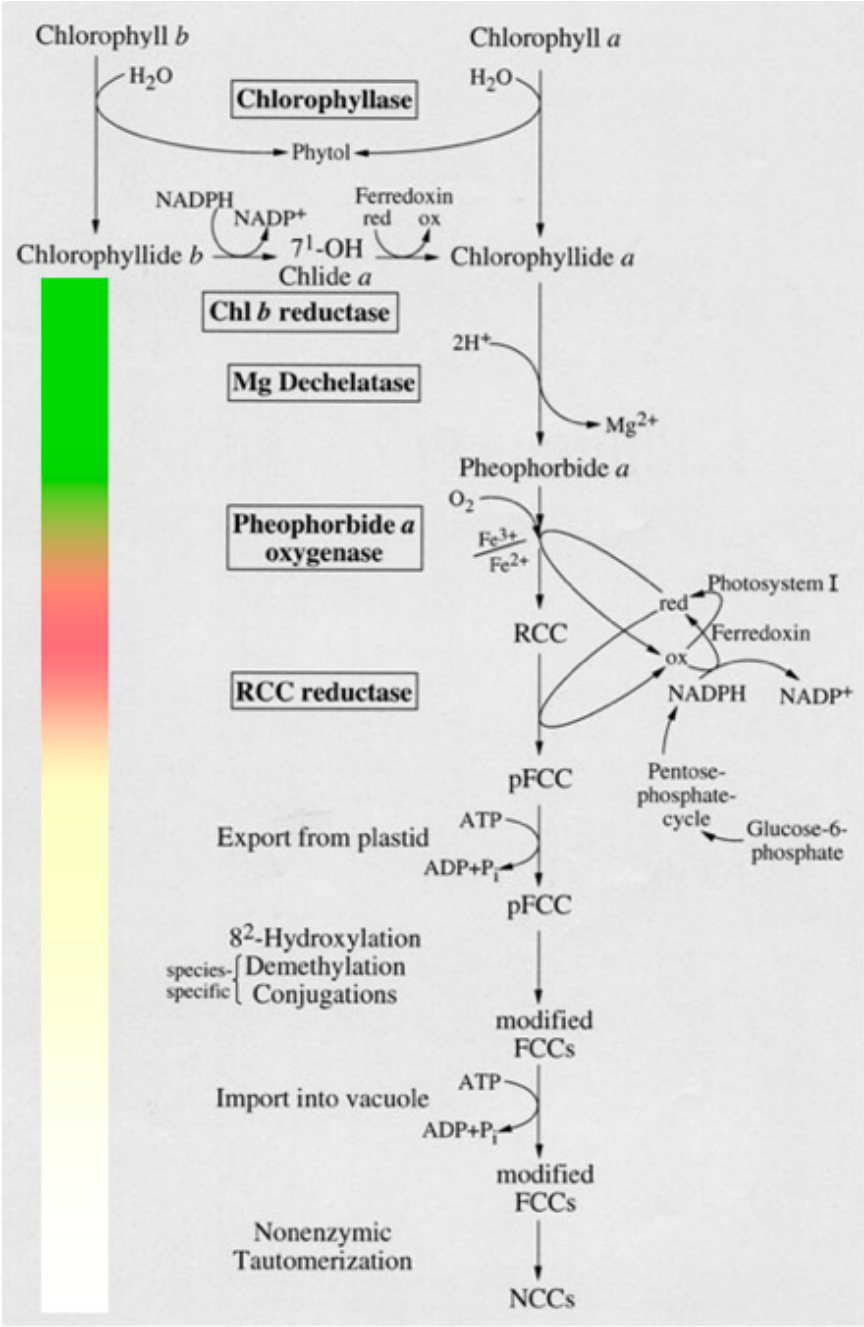
If ‘Y’ regulates ‘X’, then it stops it, starts it, slows it down, or speeds it up.

- How does a process ‘Y’ stop, start, speed up or slow down a process?
  - 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?

The future: F-P links to the rescue!

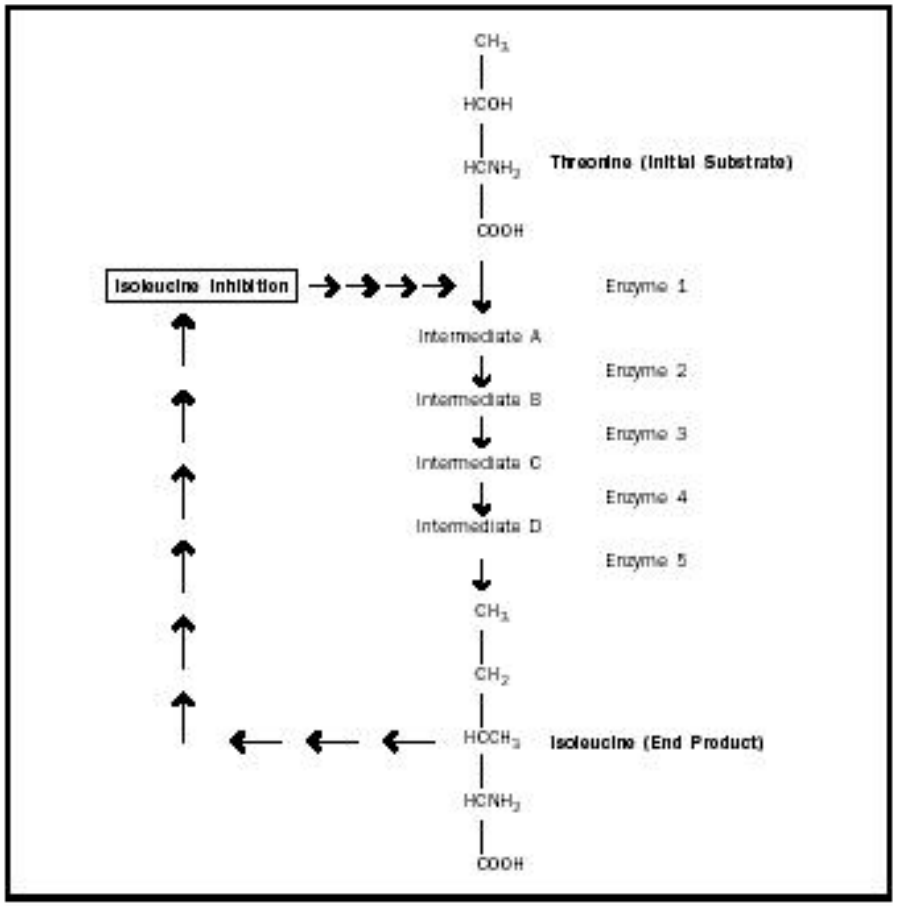
- As I told you earlier in the meeting, we plan to link all molecular functions with biological processes.
  - If we use this model, then all biological processes can be thought of as a series of molecular functions; one that starts it, one or more that are the middle, and one that stops it.
  - If another process modulates the activity of one of those molecular functions, then it regulates the process.
- BUT, the definition of the process is subjective
  - GO needs to reflect the community consensus about what functions are part of a process and what functions are not.

Sometimes
it will be easy



Sometimes
it will be impossible

Sometimes ‘Y’ might be both!



In this case, we could say that the last enzyme activity in the pathway *negatively_regulates* the pathway and is *part_of* the pathway

Even for the easy ones

- It is a huge amount of work for ontology developers to define processes based on their molecular functions
  - Different pathways/different organisms
  - Need for process-specific functions, remember why molecular function was separate from biological process in the first place!
- BUT GO HAS ALWAYS EVOLVED THIS WAY!

What's an annotator to do?

- Use your biological knowledge
  - If it is a well-known pathway and it hasn't been fully represented in GO (so far that's everything), then use our knowledge to decide if the function is part of the pathway, regulates the pathway or does both.
- What if nobody knows and it's confusing?
  - Do the best you can to reflect what the paper is saying. Presumably the author will give us some hints about what is really going on and they probably know more than we do.
  - If there is still doubt, do what we have always done in the past and annotate to the less specific parent. (Lots of annotations were made when 'regulates' was 'part_of' and no one thought they were confused.)
  - Some day, we will be able to use the ontology and the set of annotations to make inferences about what is really going on.