

Author: Chris Mungall
Contributors: <add names here>
Type: White Paper
Status: Request for comments
Projects: GO
Audience: curators, ontologists
Keywords: cellular processes
Created: 2016
Updated: 2016
Repo:
See also:

See also: [my slides from Corvalis](#) 28-37

And gist: <https://gist.github.com/cmungall/4ed28123c3db832a7d99cbdd8e8a5920>

Continued from:

<https://github.com/geneontology/go-ontology/issues/12796>

Issues with cellular-X terms:

- it's lopsided. i.e. we don't care to make parallel supracellular-X terms
- term inflation -> increased number of hypotheses in enrichment analyses -> artificially increased p values
- Inconsistent design pattern - or worse, many not axiomatized at all
- No clear guidelines on when a cellular X is to be created
- What's the use case?
- Has a user ever wanted to distinguish these?
- undocumented hidden GCIs (see appendix); are these arbitrary

What are the use cases for the grouping 'cellular process'?

PAINT has special behavior for cellular processes, these are visually grouped ahead of other BPs. The implicit assumption is that this is a useful way to filter out BPs that are less valid for propagation. ***This implicit assumption may not align with a literal OWL definition of 'cellular process'.***

PomBase uses it to exclude processes that shouldn't apply to most or all Pombe GPs.

Here's an example of the issue from PAINT, where because of this distinction two terms (cellular response to oxidative stress and response to oxidative stress) with a direct parent-child relationship are split apart (column 1 and the column with the vertical yellow lines)

The screenshot shows the PAINT interface with a search for 'oxid' yielding 21 matches. The results list includes terms like 'superoxide dismutase activity', 'response to oxidative stress', and 'cellular response to oxidative stress'. The 'Annotation Matrix' shows a grid of annotations for various proteins. The bottom table shows annotations for the term 'zinc ion binding' (GO:0008270) and 'removal of superoxide radicals' (GO:0019430), with references to PAINT_REF:10003 and various database IDs like HGNC:11179 and ENSEMBL:ENSBTAG0000018854.

What are the implications of a mass merge?

Different options here. One radical solution is to place metabolism directly under cellular process. The objection is that some metabolic processes may involve multiple cells, possibly far apart. But I would argue that the 'cellular' prefix is more about the lens through which you are looking at the process. You can make LEGO model of it. Metabolism can be looked at through a cellular lens; crudely this might mean that you can describe from start to end using a tractable number of MF instances. But for something like development, would want to take a view from further up.

I think we need to sketch out what could reasonably included under metabolism on an organism level. For many chemicals, key metabolic processes occur only in the liver - other processes elsewhere. Presumably we don't want the transport terms under a term for organismal level metabolism, but if we consider a process that occurs partly in some peripheral cell type and partly in the liver to be a metabolic process, then presumably this is multicellular or organism level metabolism. What about metabolic processes occurring in the blood? (assume there are some).

Some ideas here:

<https://gist.github.com/cmungall/4ed28123c3db832a7d99cbdd8e8a5920>

Defining cellular process

Possible tight, unambiguous definitions of cellular:

1. **'Cellular process'**: A process occurring in exactly one cell. Cell in this context includes cell surface and encapsulating structures belonging to the cell in question, such as a cell wall.
 - EquivalentTo: BP that occurs in some GO:cell ?

This would exclude some classes currently under cellular (e.g. cell-cell communication; organization of ECM components), but it gets rid of some of much of the ambiguity. The cell surface clause is meant to allow for processes that start on or occur at the cell surface, e.g. signal transduction).

2. 'Cellular process': Any process that occurs in or is mediated by a single cell. See

APPENDIX

My very hacky way to probe for finding hidden cellular GCIs:

```
$ obo-grep.pl -r 'name: cellular ' gene_ontology_write.obo | obo-grep.pl -r intersection_of -|
obo-grep.pl --neg -r 'of: .* cellular ' - | obo-grep.pl --neg -r 'occurs_in CL:0000000' -|
obo-filter-tags.pl -t name -t intersection_of -
```

[Term]

name: cellular aldehyde metabolic process

intersection_of: GO:0008152 ! metabolic process

intersection_of: has_participant CHEBI:17478 ! aldehyde

*# this is a hidden GCI as it encodes a rule that ANY metabolic process with aldehyde is cellular.
I.e. (according to current way of thinking) it is never organism level. Is this true? How do we*

know for sure? Is this an accidental hidden GCI or a deliberate one? If the latter, where is the evidence?

similar examples follow...

[Term]

name: cellular amino acid metabolic process

intersection_of: GO:0008152 ! metabolic process

intersection_of: has_participant CHEBI:37022 ! amino-acid anion

[Term]

name: cellular aromatic compound metabolic process

intersection_of: GO:0008152 ! metabolic process

intersection_of: has_participant CHEBI:33655 ! aromatic compound

[Term]

name: cellular amino acid biosynthetic process

intersection_of: GO:0009058 ! biosynthetic process

intersection_of: has_output CHEBI:37022 ! amino-acid anion

[Term]

name: cellular amino acid catabolic process

intersection_of: GO:0009056 ! catabolic process

intersection_of: has_input CHEBI:37022 ! amino-acid anion

[Term]

name: cellular component morphogenesis

intersection_of: GO:0009653 ! anatomical structure morphogenesis

intersection_of: results_in_morphogenesis_of GO:0005575 ! cellular_component

[Term]

name: cellular ketone metabolic process

intersection_of: GO:0008152 ! metabolic process

intersection_of: has_participant CHEBI:17087 ! ketone

[Term]

name: cellular modified amino acid catabolic process

intersection_of: GO:0009056 ! catabolic process

intersection_of: has_input CHEBI:83821 ! amino acid derivative

[Term]

name: cellular modified amino acid biosynthetic process

intersection_of: GO:0009058 ! biosynthetic process

intersection_of: has_output CHEBI:83821 ! amino acid derivative

[Term]

name: cellular alkane metabolic process

intersection_of: GO:0008152 ! metabolic process

intersection_of: has_participant CHEBI:18310 ! alkane

[Term]

name: cellular alkene metabolic process

intersection_of: GO:0008152 ! metabolic process

intersection_of: has_participant CHEBI:32878 ! alkene

[Term]

name: cellular alkyne metabolic process

intersection_of: GO:0008152 ! metabolic process

intersection_of: has_participant CHEBI:22339 ! alkyne

[Term]

name: cellular amide metabolic process

intersection_of: GO:0008152 ! metabolic process

intersection_of: has_participant CHEBI:32988 ! amide

[Term]

name: cellular modified histidine metabolic process

intersection_of: GO:0008152 ! metabolic process

intersection_of: has_participant CHEBI:24599 ! histidine derivative

[Term]

name: cellular modified histidine catabolic process

intersection_of: GO:0009056 ! catabolic process

intersection_of: has_input CHEBI:24599 ! histidine derivative

[Term]

name: cellular modified histidine biosynthetic process

intersection_of: GO:0009058 ! biosynthetic process

intersection_of: has_output CHEBI:24599 ! histidine derivative