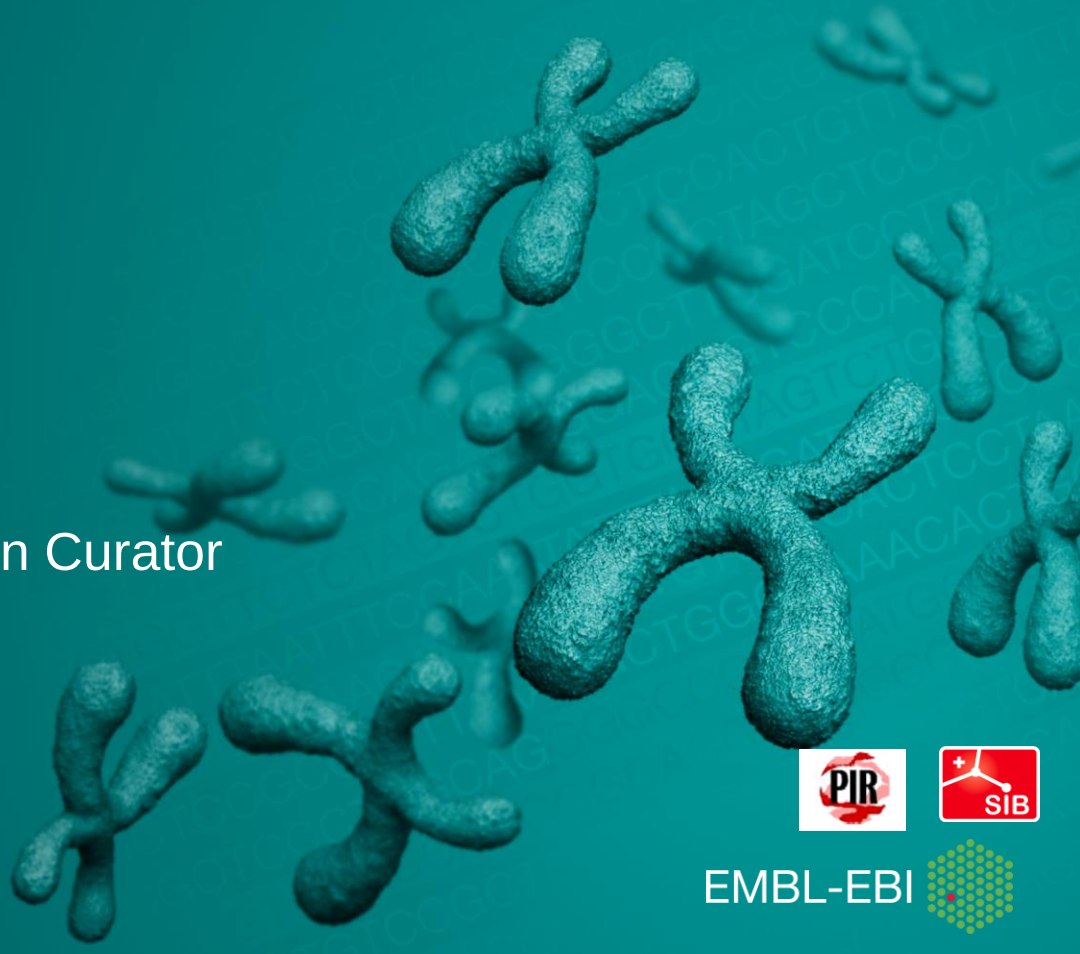
Use of Gene Ontology Annotation to understand the peroxisome proteome in humans



Prudence Mutowo

UniProt-Gene Ontology Annotation Curator

EMBL-EBI

UniProt

PIR

SIB

EMBL-EBI

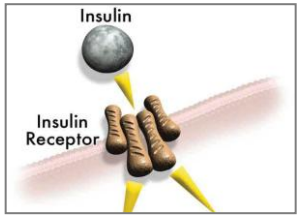
Manual annotation in UniProt-GOA

Literature-based GO annotation to UniProt accessions

GO terms aim to describe the ‘normal’ functions/ processes/locations of gene products

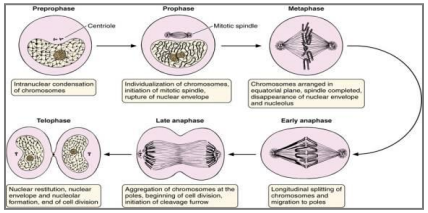
Molecular Function

e.g. insulin receptor activity



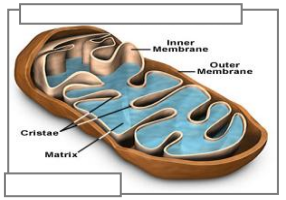
Biological Process

e.g. cell cycle



Cellular Component

e.g. mitochondrion



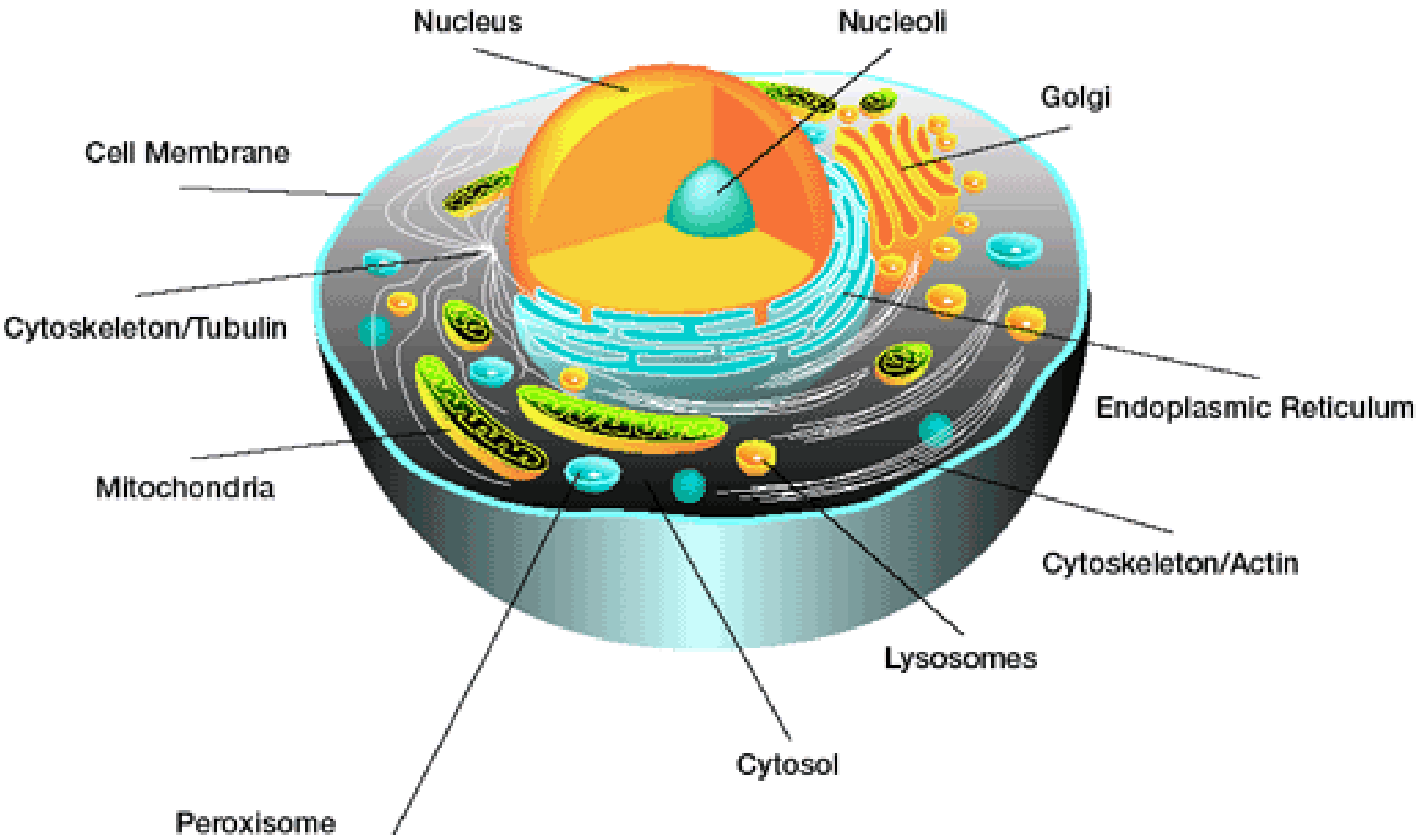
NO: pathological processes, experimental conditions or temporal information

Choosing a protein set

Manual curation expensive, time consuming, slow

- What considerations when selecting protein sets for manual curation?
-
- What are the benefits to the data set?

The human peroxisome



Organelle-centric annotation: The human peroxisome

Peroxisome functions

- Hydrogen peroxide metabolism
- Detoxification
- Fatty acid metabolism

1. Involvement in human health and disease. (Refsum, Zellweger)

2. Dedicated research efforts
****Integrated project to decipher the biological function of peroxisomes in health and disease (finished in Dec 2008)**

Reasons for curation focus

1. Area of biological focus (databases, active community of researchers , consortia etc)

2. Interesting science- involvement in human health and disease

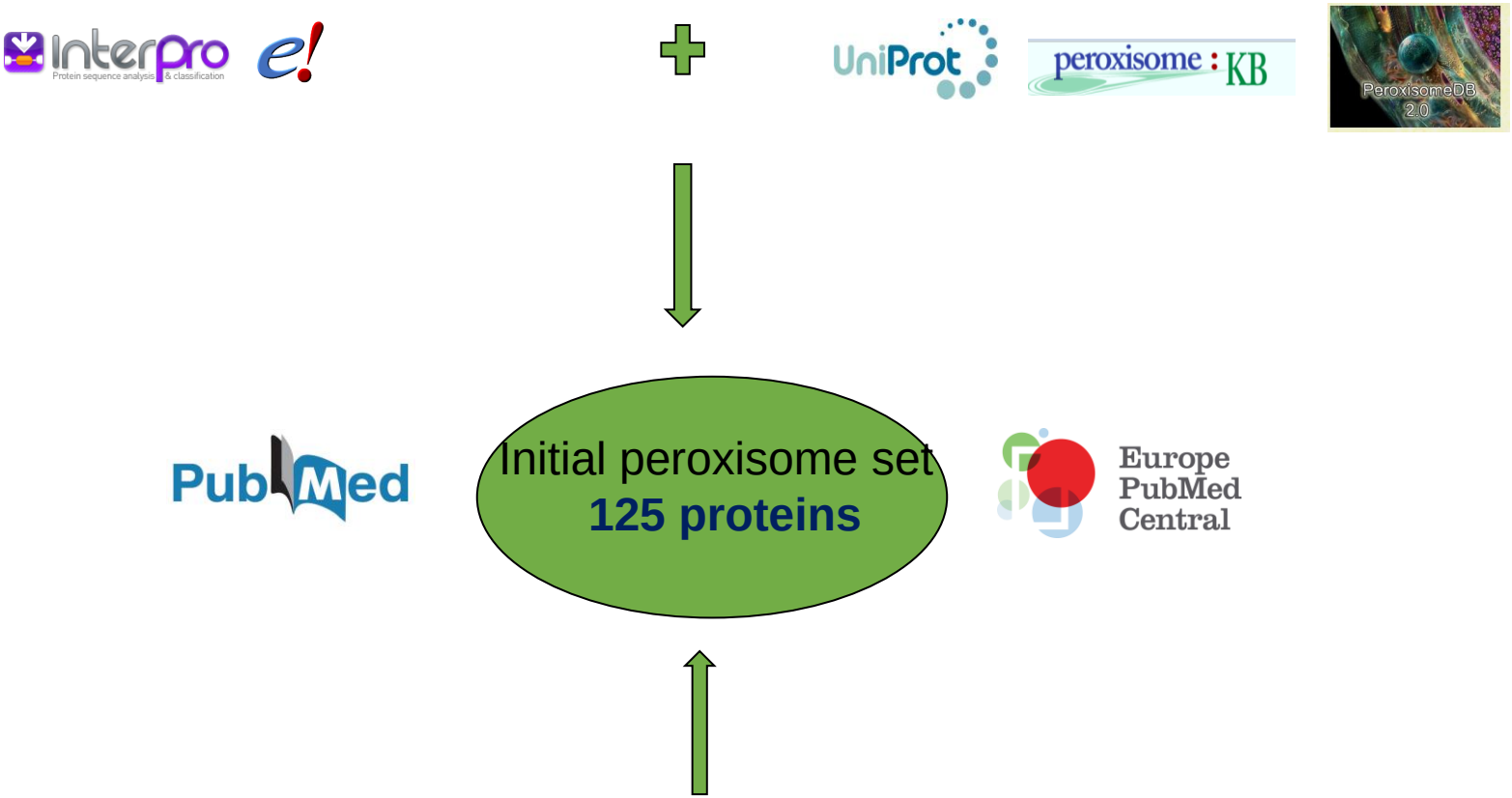
3. Protein set is a concise manageable number (>200 starting set)

4. Availability of a wealth of literature to curate

Organelle focused GO annotation: Methodology

Predictive methods

Peroxisome resources



Full and comprehensive GO annotation

Identification of proteins with experimental peroxisomal localisation

Outcomes of the focused annotation effort

88 human peroxisomal proteins fully curated

- Proteins without prior direct evidence for peroxisome location annotations captured e.g insulin degrading enzyme

296 other non-peroxisomal proteins also annotated during the process

- Demonstrating benefits of annotation policy that encourages curating papers fully

1551 annotations created to all proteins

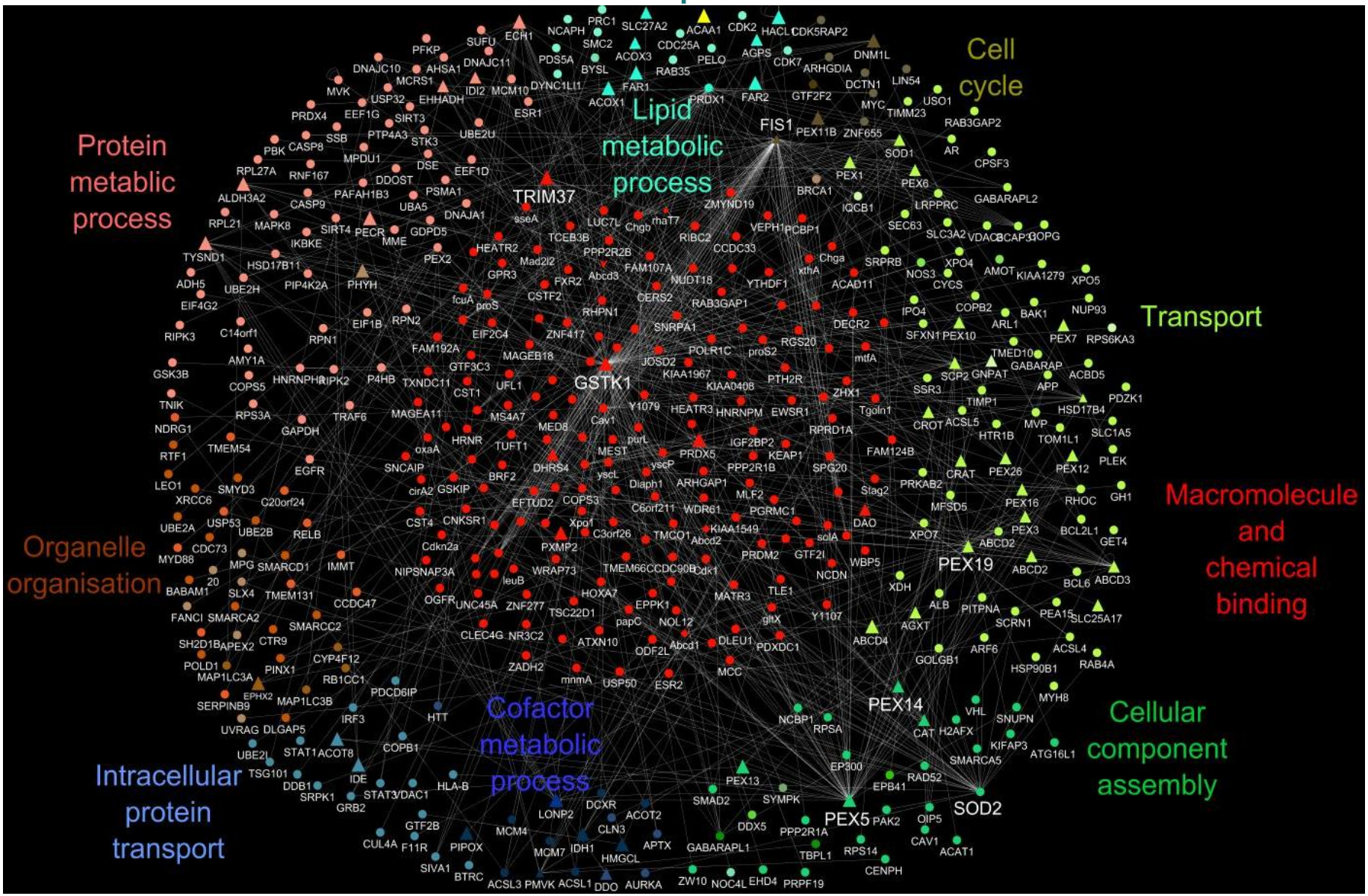
38 new peroxisome related GO terms requested

- Growth in the ontology

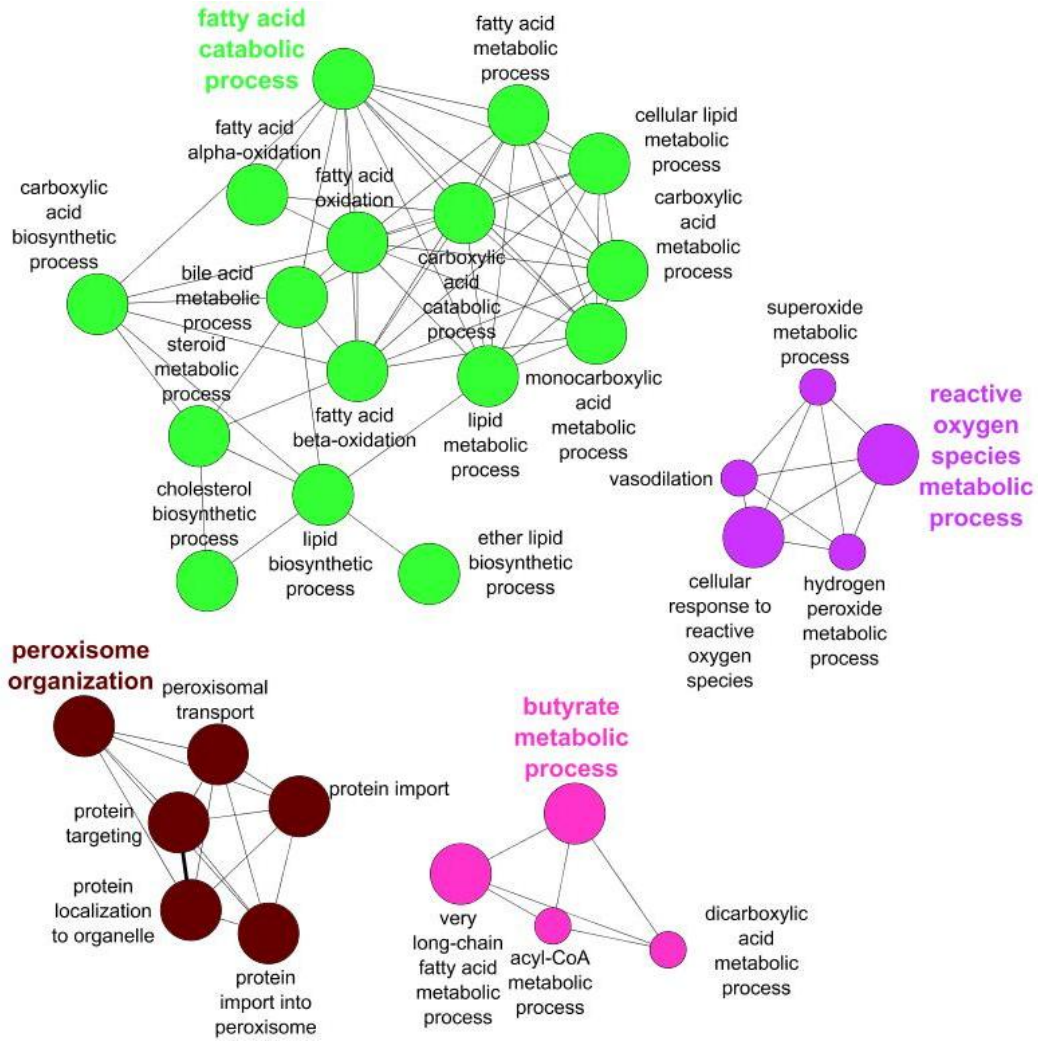
Use case for the annotation extension in GO

contextual information captured for activities and location
e.g cell and tissue type
isoform/feature chain
localisation dependency
e.g majority of peroxisomal proteins traditionally described in liver, kidney, new tissues and cell types including brain, testes

GO slims to obtain overview of peroxisome interactome functions



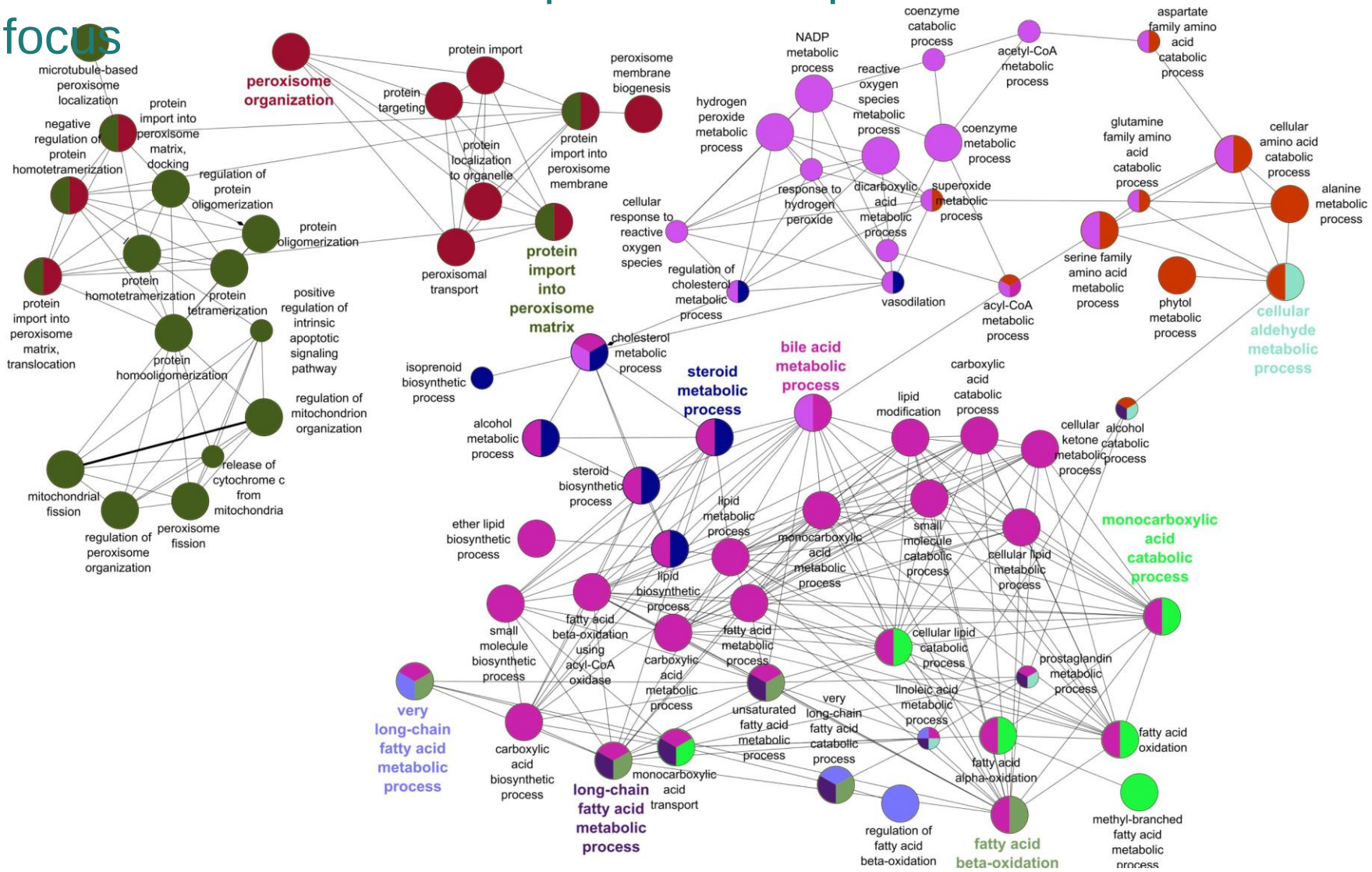
GO enrichment human peroxisomal proteins, before annotation focus



Gene ontology enrichment of 88 human peroxisome proteins before the focused manual peroxisome protein annotation effort using the June 2010 version of the ontology

Analysis done using ClueGO on Cytoscape visualisation platform

GO enrichment: Human peroxisomal proteins after annotation focus



Gene ontology enrichment of 88 human peroxisome proteins after the focused annotation effort

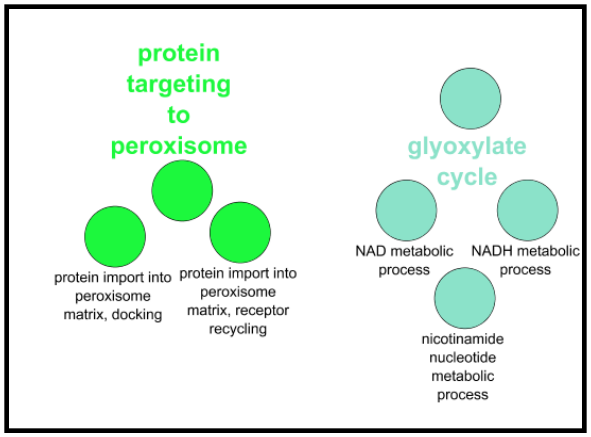
Main benefits to dataset: Term enrichment

- Increase in the number of enrichment terms in Figure 2 compared to Figure 3.
- New processes have also been enriched in the second analysis.
  Examples include the alcohol metabolic processes, peroxisome fission, carboxylic acid biosynthetic processes
- Added depth and specificity has also been added to the protein set

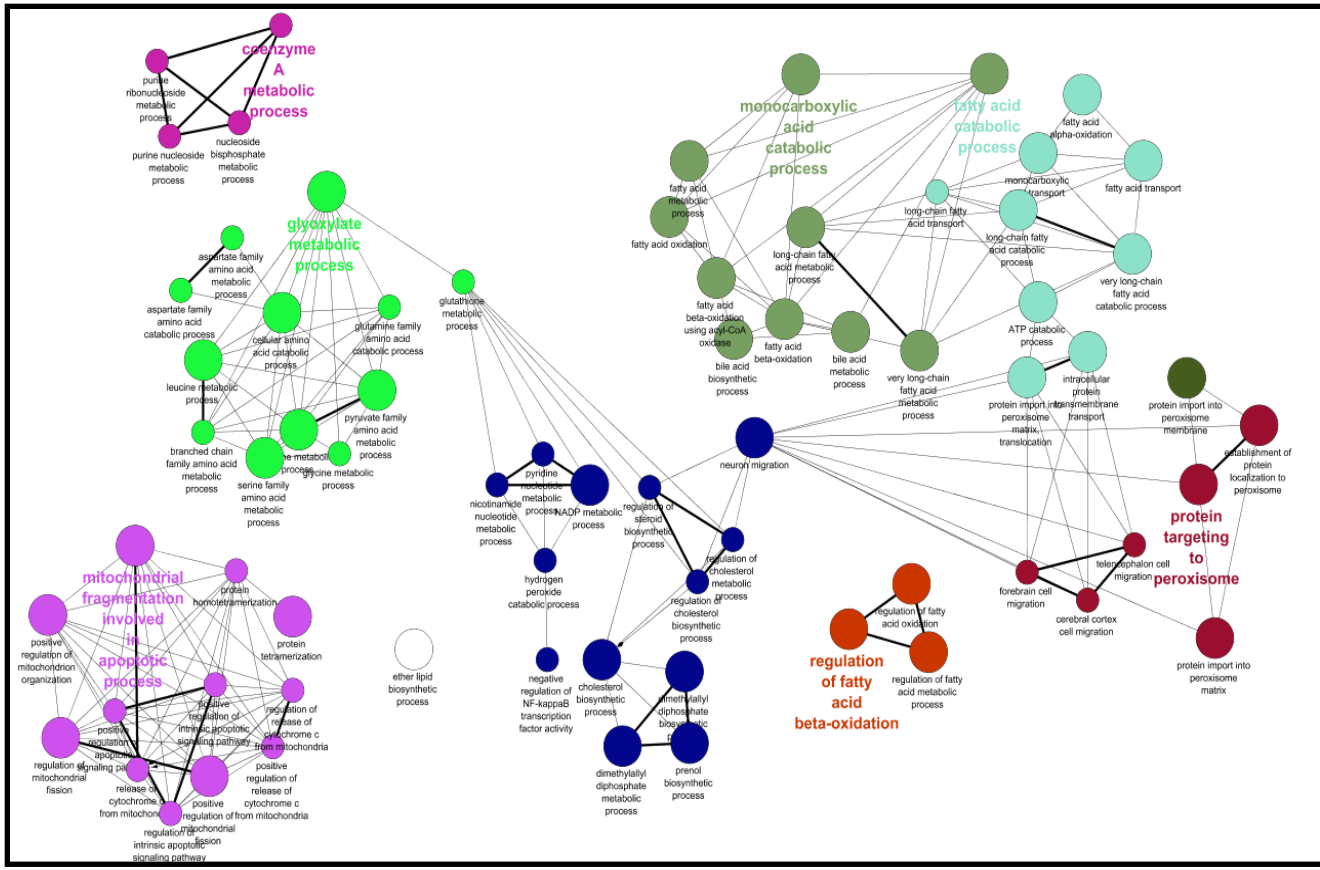
  e.g fatty acid metabolism more granular terms like fatty acid beta oxidation using acyl CoA oxidase being added thus giving better depth of information to the peroxisomal fatty acid oxidation process

GO enrichment species comparison: S. cerevisae and human peroxisome proteins

Yeast enriched



Human enriched



Comparison of GO terms enriched between yeast and human
Glyoxylate cycle
Fatty acid beta oxidation
Mitochondrial and apoptotic processes

Observations from yeast vs human comparison

Species differences

- forebrain cell migration , neuron cell migrations expected only human set

Differences due to different organelle functions

Human

- involvement in apoptosis
- mitochondrion organization

Yeast

- glyoxylate cycle occurs in yeast and human equivalent occurs in the mitochondrion

Caveats

Experiments may yet to be done for some differences

Literature not yet curated

Some differences in curational priorities/style between different groups

Overall benefits of the organelle-focused annotation strategy

- provide more in-depth information for a set of proteins
- allow a global GO-centric overview for an organelle's function
- permit species comparison for similarly curated organisms
- highlights the importance of conducting GO annotation using same/similar standards with groups curating other organisms
- Allows curators to focus on a specific field and use any background reading to full benefit

Applications for work carried out in this way

Potential for focused work to be useful to researchers as a baseline determinant of the state of literature for related projects

Potential for GO annotation to be an integral part of the new initiative

Encourages sustained communication with the scientific community in a particular area of knowledge

Acknowledgments

UniProt Team Leaders: Claire O'Donovan and Maria Jesus Martin

UniProt curators and programmers

UniProt-GOA project

R. Huntley

Y. Alam-Faruque

T. Sawford

P. Mutowo

UniProt Consortium, GO Consortium , The IntAct Team

Funding

National Institutes of Health (NIH)

EMBL

Kidney research UK (KRUK)

British Heart Foundation (BHF)

Opportunities

Ontology developer position with the GO office (http://www.ebi.ac.uk/about/jobs)

UniProt

PIR

SIB

EMBL-EBI