09:00-12:00

- Interpretation and visualization of CAFE5 results
- Introduction to GO enrichment analyses
- GO enrichment analyses
- Wrap-up discussion

12:00-13:00 Lunch

13:00-14:00

- Talk by Alex Suh

14:00-14:15 Break

14:15-16:00

- Analyses of repeats in R

Genome assembly, annotation and comparative genomics

Day 3, morning

Teachers: Lars Grønvold, Thu-Hien To, Bram Danneels, Helle Tessand Baalsrud, Ole K. Tørresen

Norwegian Biodiversity & Genomics Conference 2024
10th April

Introduction to GO enrichment analyses (with custom annotations)

Outline:

- Theory
  - About the gene ontology
  - Enrichment analysis and g:Profiler
- Practical exercise - GO enrichment of expanded/contracted gene families
  - (on the git repo)
    - Extracting GO terms from GFF files
    - Use orthofinder Hierarchical Orthogroups (HOG) to combine GO terms
    - Using GMT helper to convert GO table to valid GMT file
    - Get list of significant gene families
    - Enrichment analysis with g:Profiler
  - Interpretation of results

```
MTKSHSEEVIVPEFNSSAKELPRPLAEKCPSIIKKFISAYDAKPDFVARSPGRVNLIGEH
IDYCDFSVLPLAIDFDMLCAVKVLNEKNPSITLINADPKFAQRKFDLPLDGSYVTIDPSV
SDWSNYFKCGLHVAHSFLKKLAPERFASAPLAGLQVFCEGDVPTGSGLSSSAAFICAVAL
AVVKANMGPGYHMSKQNLMRITVVAEHYVGVNNGGMDQAASVCGEEDHALYVEFKPQLKA
TPFKFPQLKNHEISFVIANTLVVSNKFETAPTNYNLRVVEVTTAANVLAATYGVVLLSGK
EGSSTNKGNLRDFMNVYYARYHNISTPWNGDIESGIERLTKMLVLVEESLANKKQGFSVD
DVAQSLNCSREEFTRDYLTTSPVRFQVLKLYQRAKHVYSESLRVLKAVKLMTTASFTADE
DFFKQFGALMNESQASCDKLYECSCPEIDKICSIALSNGSYGSRLTGAGWGGCTVHLVPG
GPNGNIEKVKEALANEFYKVKYPKITDAELENAIIVSKPALGSCLYEL
```

Function

Phosphorylates alpha-D-galactose to alpha-D-galactose-1-phosphate in the first step of galactose catabolism

Controlled vocabulary for function definition

- Standardized and machine readable
- Examples:
  - **Gene Ontology (GO)** – Defines concepts/classes to describe gene function and relationships between these concepts.
    - Example: GO:0004335 - galactokinase activity
  - **Enzyme Commission (EC)** - Numerical classification scheme for enzymes, based on the chemical reactions they catalyze.
    - Example: EC 2.7.1.6 - Galactokinase
  - **Kyoto Encyclopedia of Genes and Genomes (KEGG)** - a collection of databases dealing with genomes, biological pathways, diseases, drugs, and chemical substances.
    - Example: map00052 - Galactose metabolism

GENEONTOLOGY
Unifying Biology

The Three Aspects:

- **Molecular function (GO:0003674)**
  - Molecular activities of gene products. E.g. "catalytic activity" or "binding activity"
- **Biological process (GO:0008150)**
  - Pathways and larger processes made up of the activities of multiple gene products. E.g. "response to cold", "signal transduction"
- **Cellular component (GO:0005575)**
  - Where gene products are active
    E.g. "nucleus" or "ribosome"

The GO hierarchy

- Parent terms are more broad
  - (less informative)
- Child terms are more specific
  - (more informative)
- Directed acyclic graph
- **All parent terms are implied**
- There are different types of relations

QuickGO - https://www.ebi.ac.uk/QuickGO
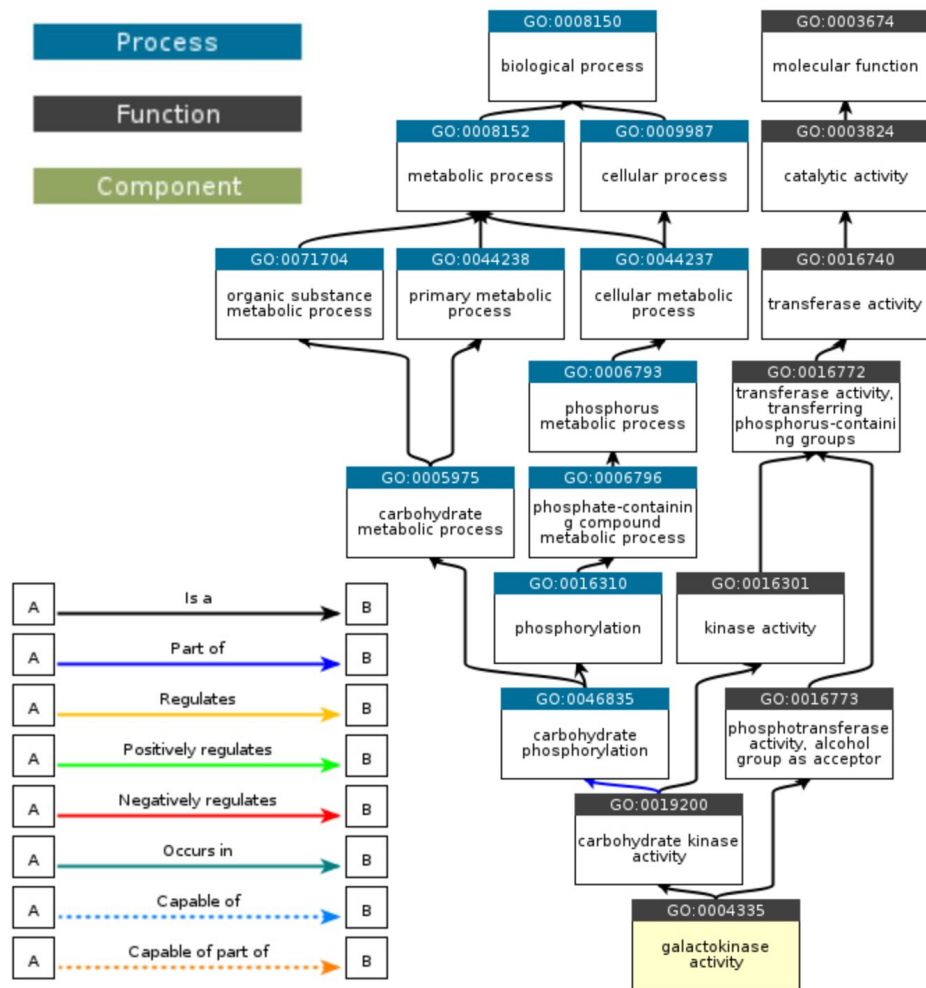
Gene Function Enrichment

- Investigate function of a list of genes (query)
- For each annotation:
  - Test if the overlap is significantly higher than expected by chance

- Obs! Correct for multiple testing!
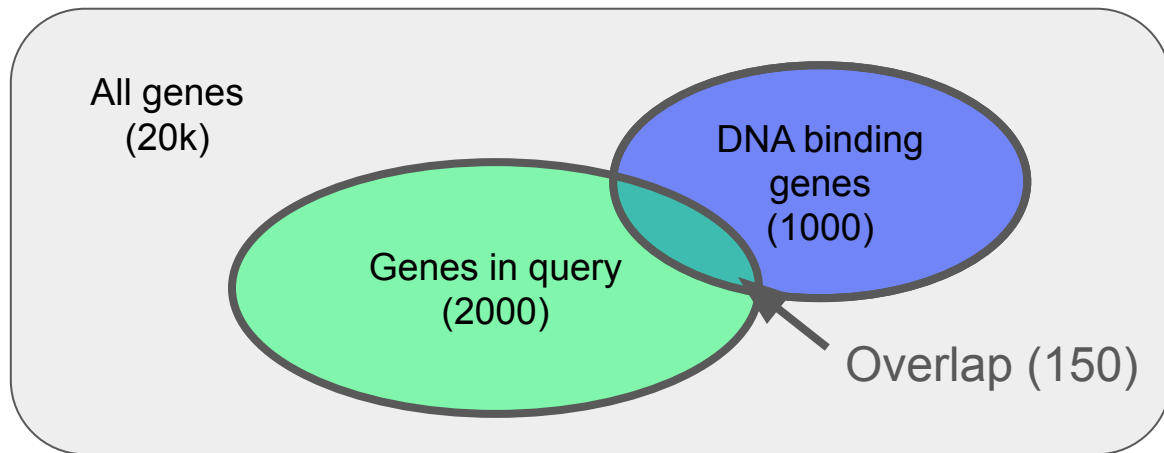
- Online functional enrichment analysis
- Works great if you have Ensembl gene IDs
- Can also automatically convert from other gene IDs (including NCBI IDs)
- **Custom annotations with GMT (GMT-helper)**
- Also has some other tools like
  - **g:Convert** Gene ID conversion (also useful to get gene names and description, or to get all genes associated with a GO)
  - **g:Orth** Ortholog lookup
- Web API with R and Python client

GMT-helper

g:Profiler

- Convert your custom GO annotations from a simple table to a GMT file compatible with g:Profiler
- Needs OBO file to propagate the terms up the GO hierarchy

TSV to GMT

Reannotate

.obo file

Practical exercise -
GO enrichment of expanded/contracted gene families

Go to:

https://github.com/ebp-nor/workshop-2024/blob/main/day3_comparative_genomics/GO-enrichment.md

Looking at the results

- Hower the dots in the plot to see the terms.
- Click to add them to the table below
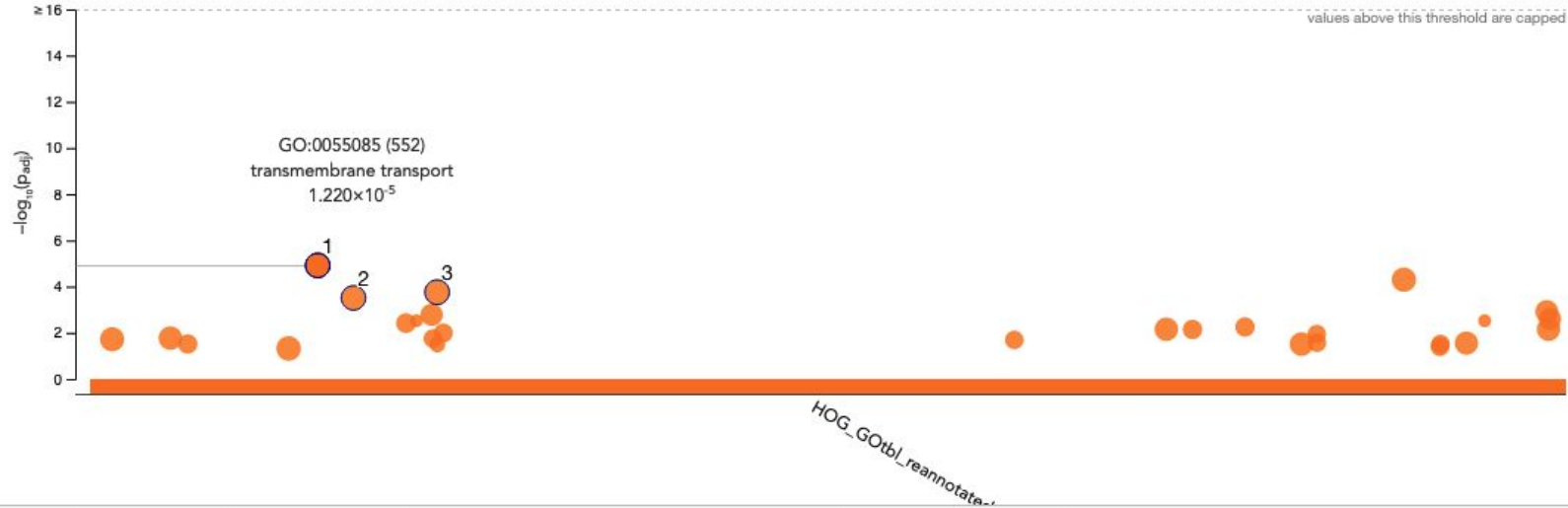
ID	Source	Term ID	Term Name	p_adj (query_1)
1	HOG_GOtb...	GO:0055085	transmembrane transport	1.220×10^{-5}
2	HOG_GOtb...	GO:0022857	transmembrane transporter activity	3.093×10^{-4}
3	HOG_GOtb...	GO:0016491	oxidoreductase activity	1.749×10^{-4}

Detailed results

Overview   Detailed Results   GO Context   Query Info

Share: Show query URL   📋   Show short link   📋

▼   Should match...   ☐ Term size   6   ⬍   ⊞⊞ ──────────   602   ⬍
                                            6 602

⬇   CSV ⚙ PNG ⚙ GEM

                                            More details hidden here

⌄ Legend

HOG_GOtbl_reannotated		stats		⟩⟩	NO.HOG0000007	NO.HOG0000010	NO.HOG0000013	NO.HOG0000020	NO.HOG0000026	NO.HOG0000028	NO.HOG0000031	NO.HOG0000032	NO.HOG0000036	NO.HOG0000043	NO.HOG0000048	NO.HOG0000049	NO.HOG0000053	NO.HOG0000059	NO.HOG0000061	NO.HOG0000065	NO.HOG0000068	NO.HOG0000069	NO.HOG0000071	NO.HOG0000073	NO.HOG0000076
☐ Term name		Term ID ✋	p_{adj}	$-\log_{10}(p_{adj})$ 0 ≤16																					
☐ transmembrane transport		GO:0055085	1.220×10^{-5}							■					■	■				■	■				
☐ transporter activity		GO:0005215	5.047×10^{-5}						■						■	■				■					
☐ oxidoreductase activity		GO:0016491	1.749×10^{-4}						■		■	■													
☐ transmembrane transporter activity		GO:0022857	3.093×10^{-4}							■					■	■				■	■				
☐ RNA polymerase II transcription regulatory region ...		GO:0000977	1.199×10^{-3}		■	■	■	■		■			■				■	■				■			
☐ RNA polymerase II cis-regulatory region sequence...		GO:0000978	1.651×10^{-3}		■	■	■	■		■			■				■					■			
☐ cis-regulatory region sequence-specific DNA bindi...		GO:0000987	2.557×10^{-3}		■	■	■	■		■			■				■					■			
☐ oxidoreductase activity, acting on a heme group of...		GO:0016675	3.013×10^{-3}																						
☐ cytochrome-c oxidase activity		GO:0004129	3.013×10^{-3}																						
☐ developmental process		GO:0032502	3.829×10^{-3}		■		■	■										■		■			■		
☐ heme binding		GO:0020037	5.642×10^{-3}							■	■														

Viewing the significant terms in QuickGO