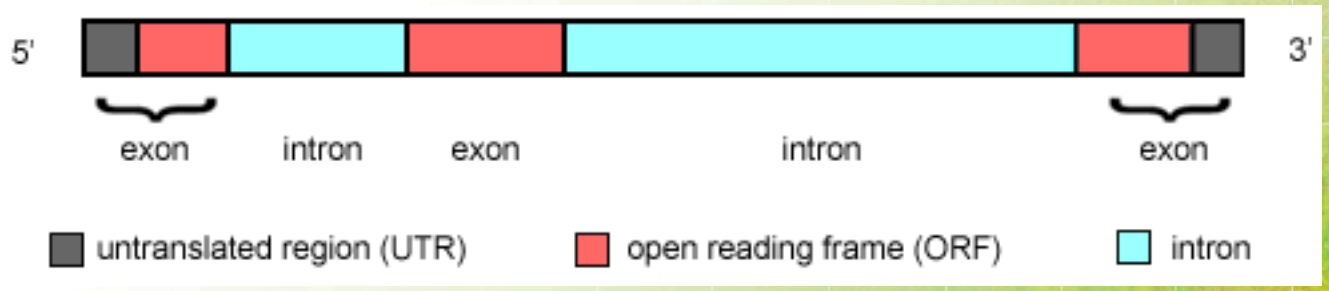
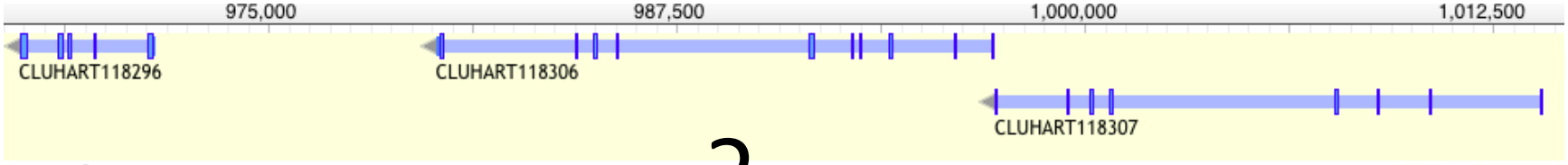
Functional annotation



Right, now we have our genes, but what do they do?



?

?

?

Insulin receptor?

Vesicle-trafficking protein?

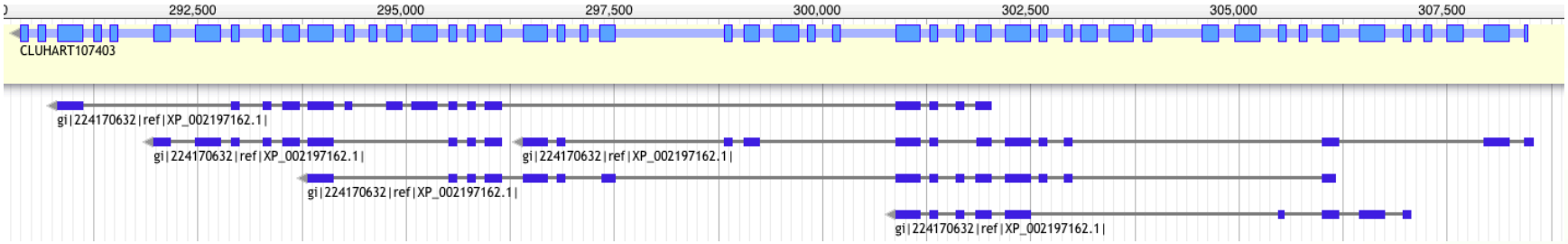
Alcohol dehydrogenase?

Aquaporin?

Transcription factor

MAP kinase kinase kinase?

But we have used proteins in our annotation!



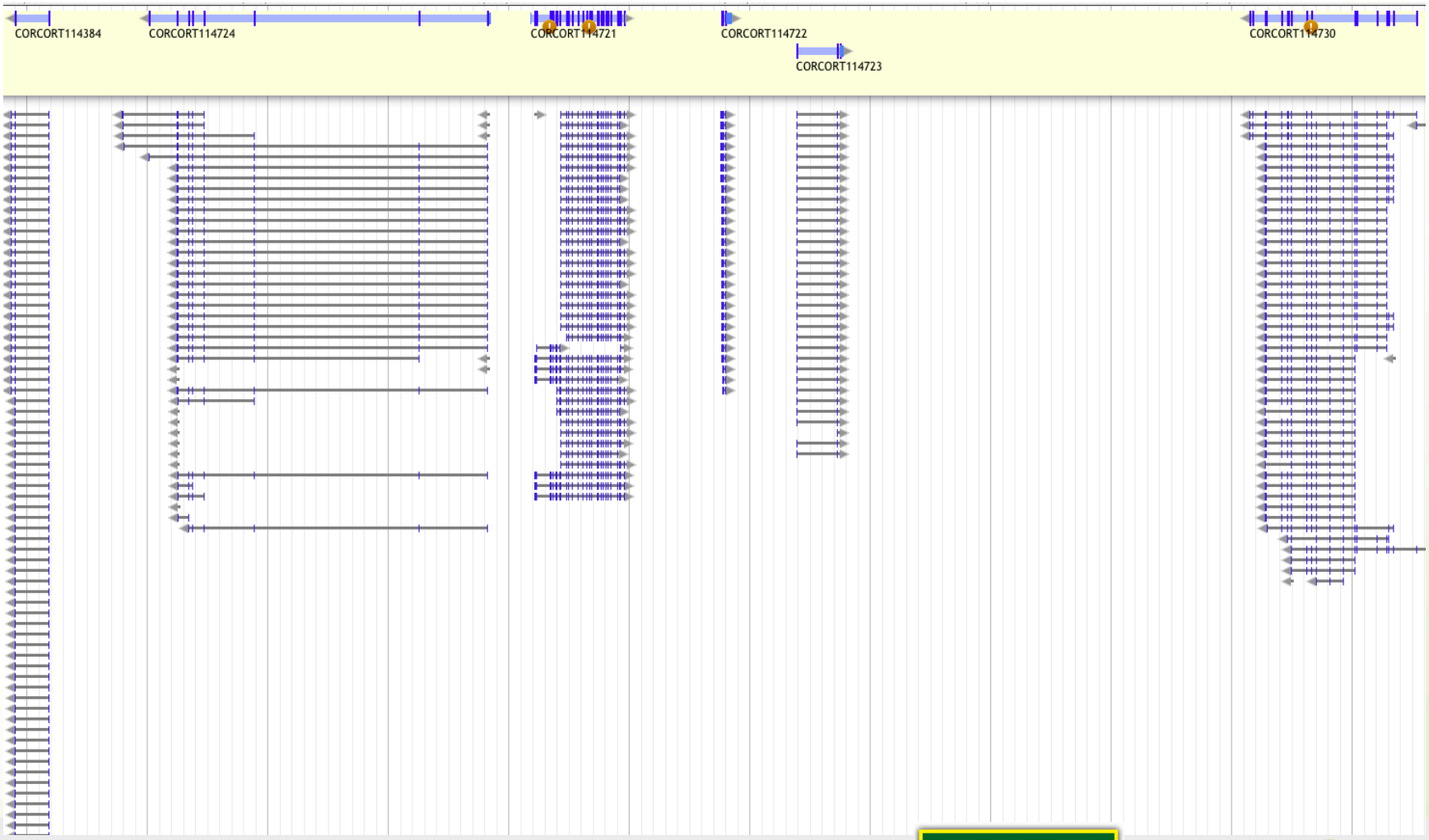
It is actually kind of complex…

… and Maker does not do this for you.

Extract sequences -> functional annotation

- Extract sequences from Webapollo or use gffread (in Cufflinks package)
- Annotate the sequences functionally using Blast and InterProScan
- Or use the (mostly) commercial alternative

Blast-based annotation - proteins



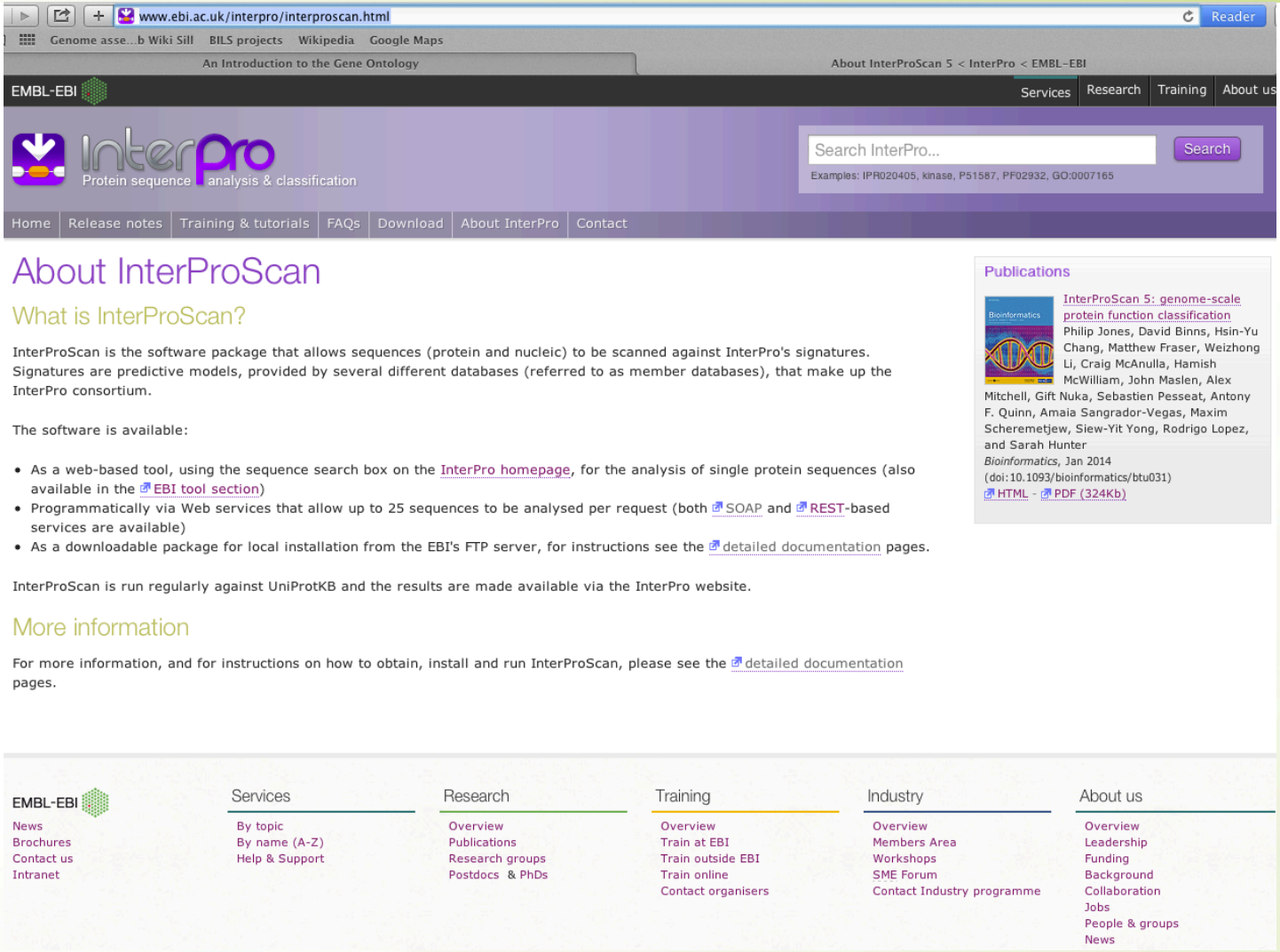
Blast-workflow

- Blast the protein-sequences from your maker-run to a protein blast-database (e.g., uniprot) using blastp from the Blast+ package
- Use Annie to extract best hits from blast-hit list and the corresponding description from uniprot-headers
- Add this information to the annotation.gff using custom-script

Blast-based approach

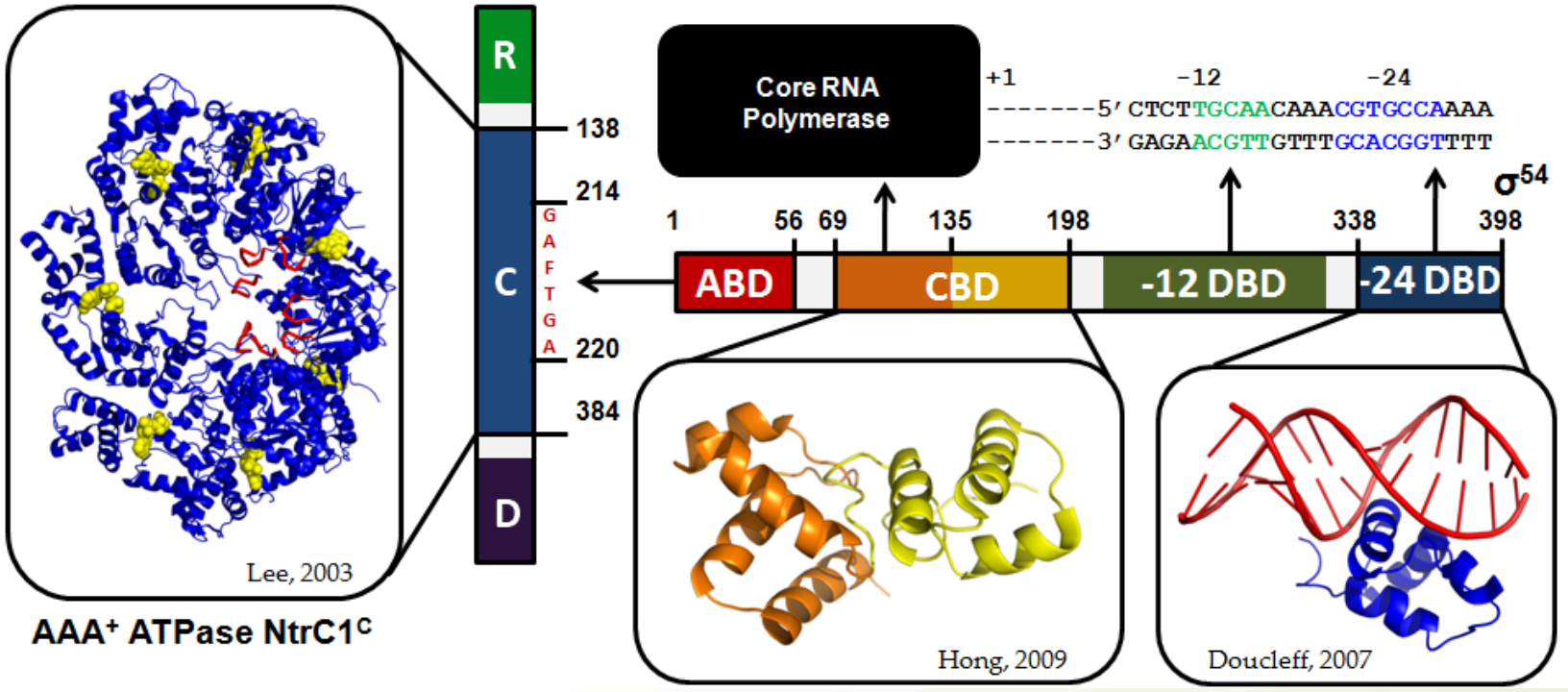
- Fairly fast and easy
- Orthology not certain - best blast-hit does not equal orthologous!

Interproscan

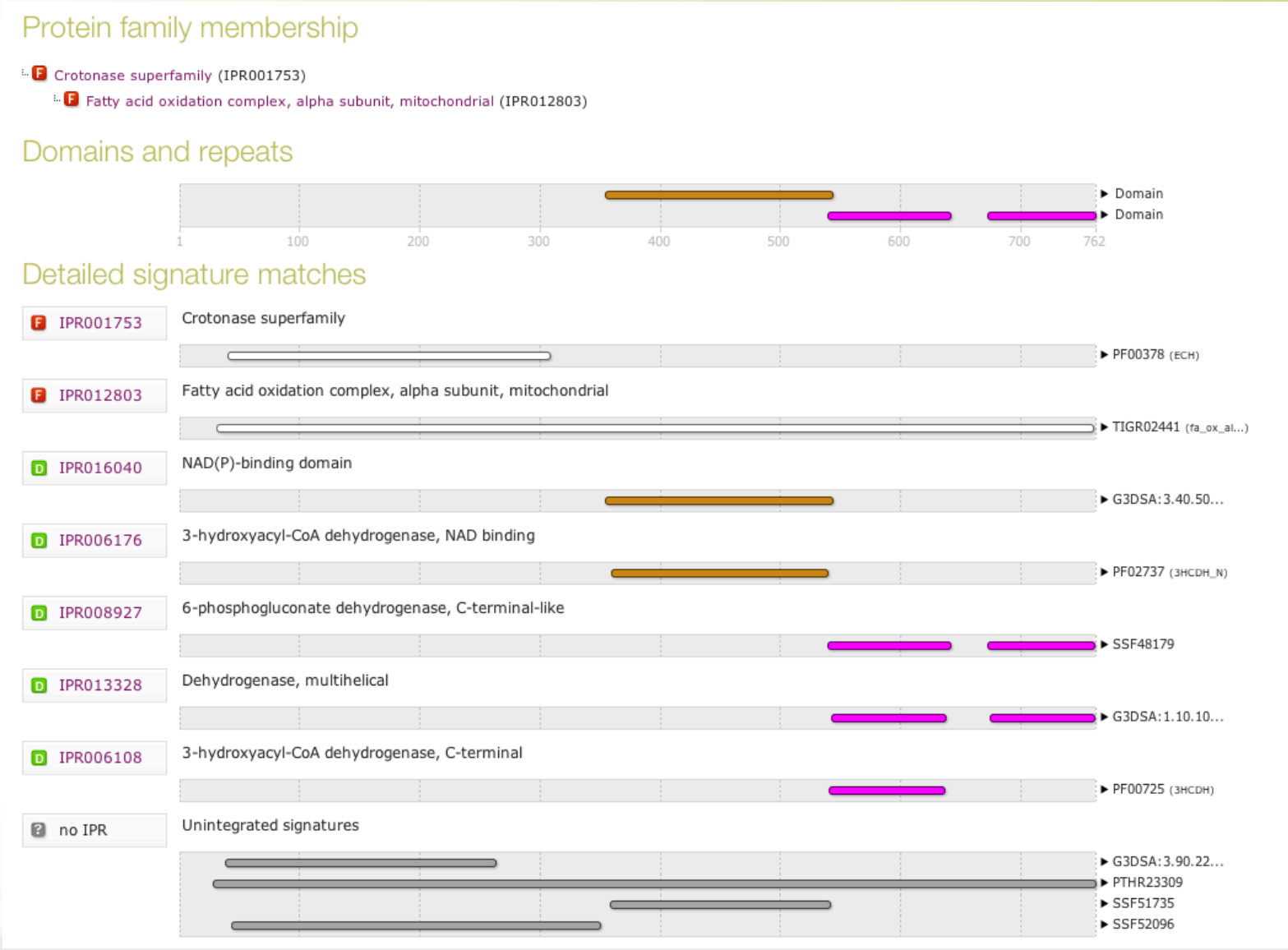


Karolinska Institutet
KTH Vetenskap och konst
Stockholms universitet
Uppsala Universitet
SciLifeLab

Sequence domains



Interproscan results



Interproscan results - GO terms

GO term prediction

Biological Process

GO:0006631 fatty acid metabolic process
GO:0006635 fatty acid beta-oxidation
GO:0008152 metabolic process
GO:0055114 oxidation-reduction process

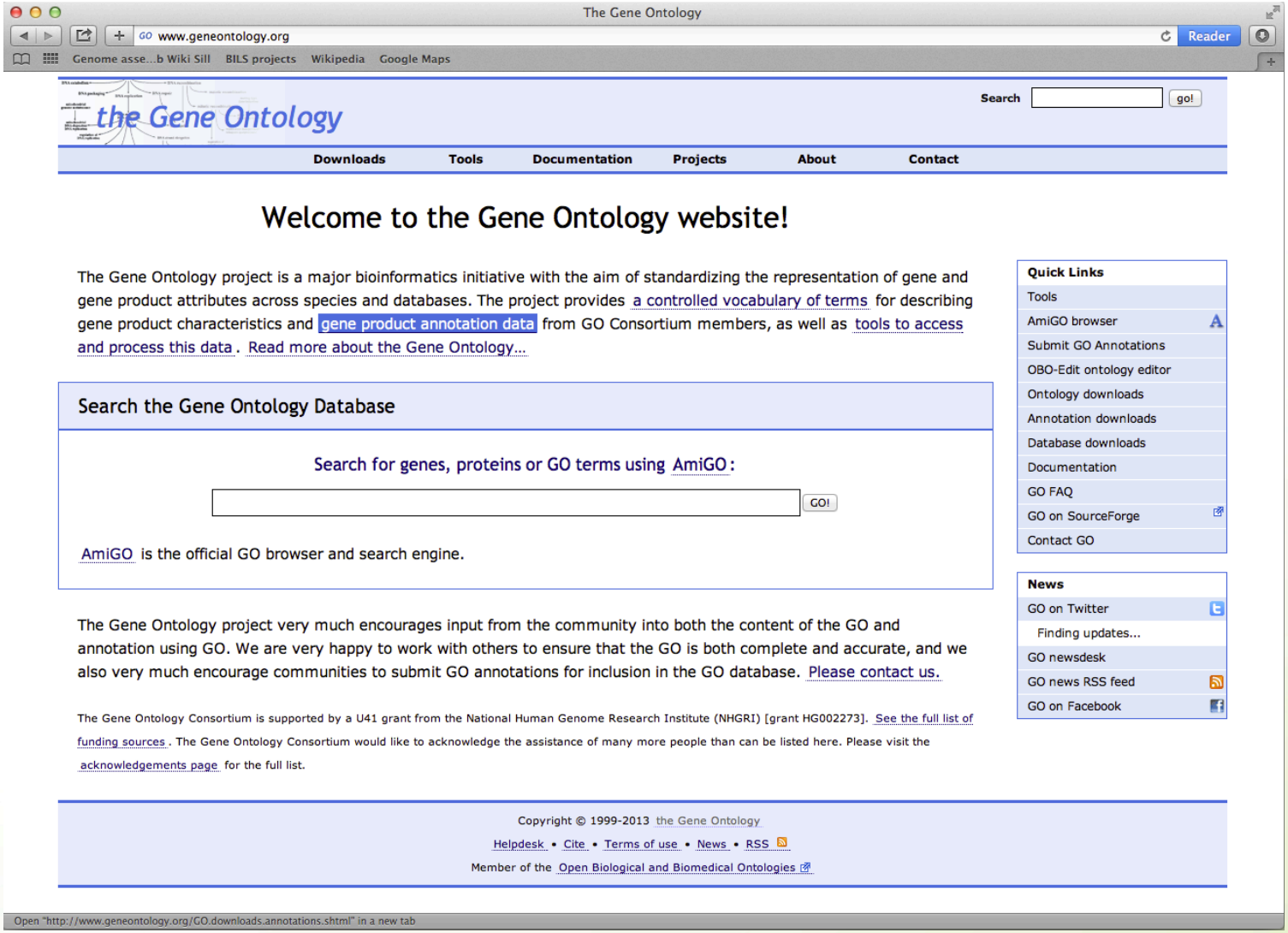
Molecular Function

GO:0003824 catalytic activity
GO:0003857 3-hydroxyacyl-CoA dehydrogenase activity
GO:0004300 enoyl-CoA hydratase activity
GO:0016491 oxidoreductase activity
GO:0016616 oxidoreductase activity, acting on the CH-OH group of donors, NAD or NADP as acceptor
GO:0050662 coenzyme binding

Cellular Component

GO:0005739 mitochondrion
GO:0016507 mitochondrial fatty acid beta-oxidation multienzyme complex

Gene Ontology



Karolinska Institutet
KTH Vetenskap och konst
Stockholms universitet
Uppsala Universitet
SciLifeLab

Gene Ontology

- C: Cellular Component, e.g., endoplasmatic reticulum, integral to plasma membrane
- P: Biological Process, e.g., pyrimidine metabolic process
- F: Molecular Function, e.g., catalytic acticivity, transporter activity

Interproscan workflow

- Run Interproscan on the protein fasta file created by maker
- Use Maker-supplied scripts to merge the interproscan-results to the Maker annotations.gff file

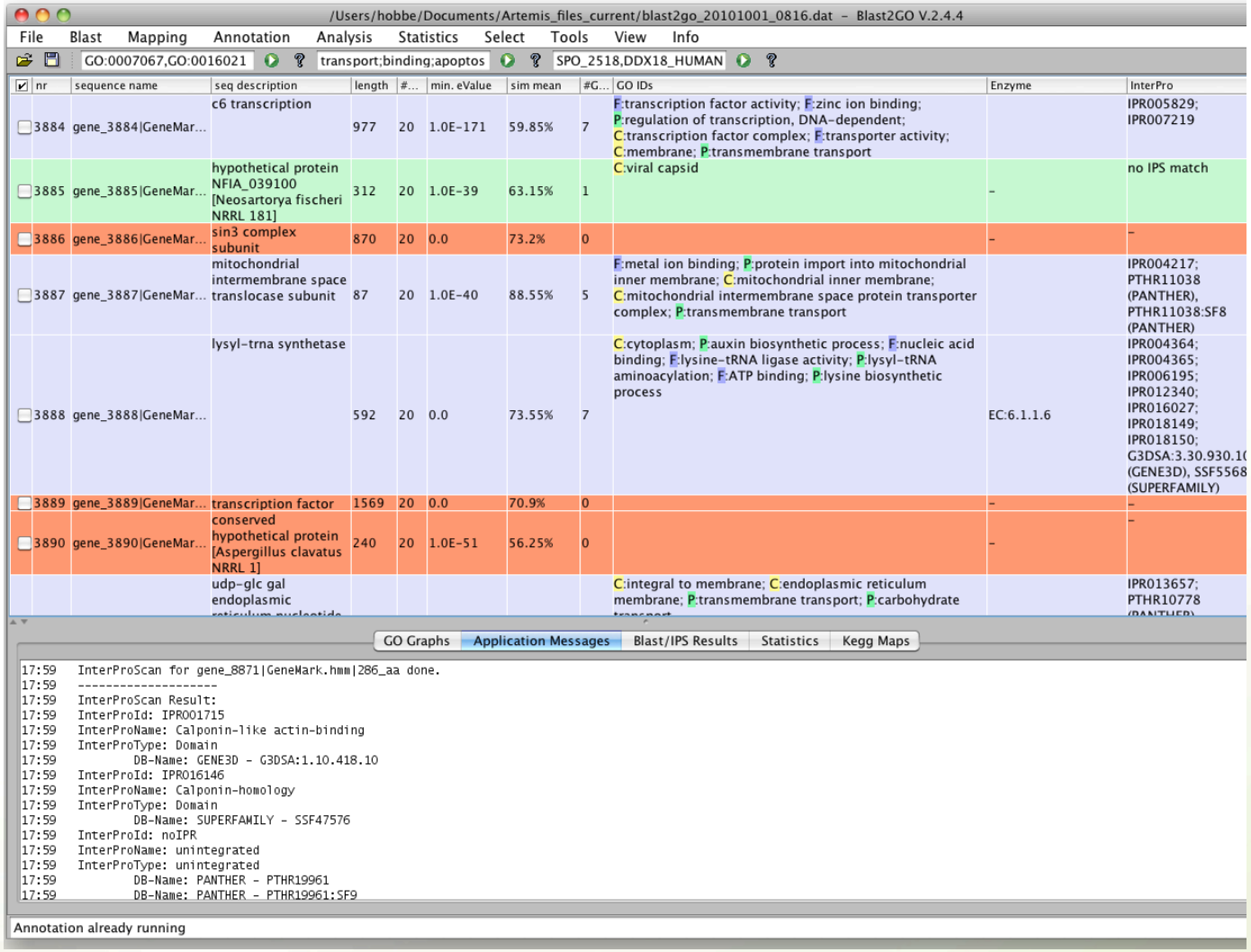
Interproscan approach

- Looks for conserved domains, so might be more reliable than blast?
- How to go from conserved domains to assigning a function for your protein?

Blast2GO

- Combines a blast-based search with a search for functional domains
- Blast at NCBI -> picks out GO terms based on blast hits and uniprot -> statistical significance test -> done!
- Interproscan

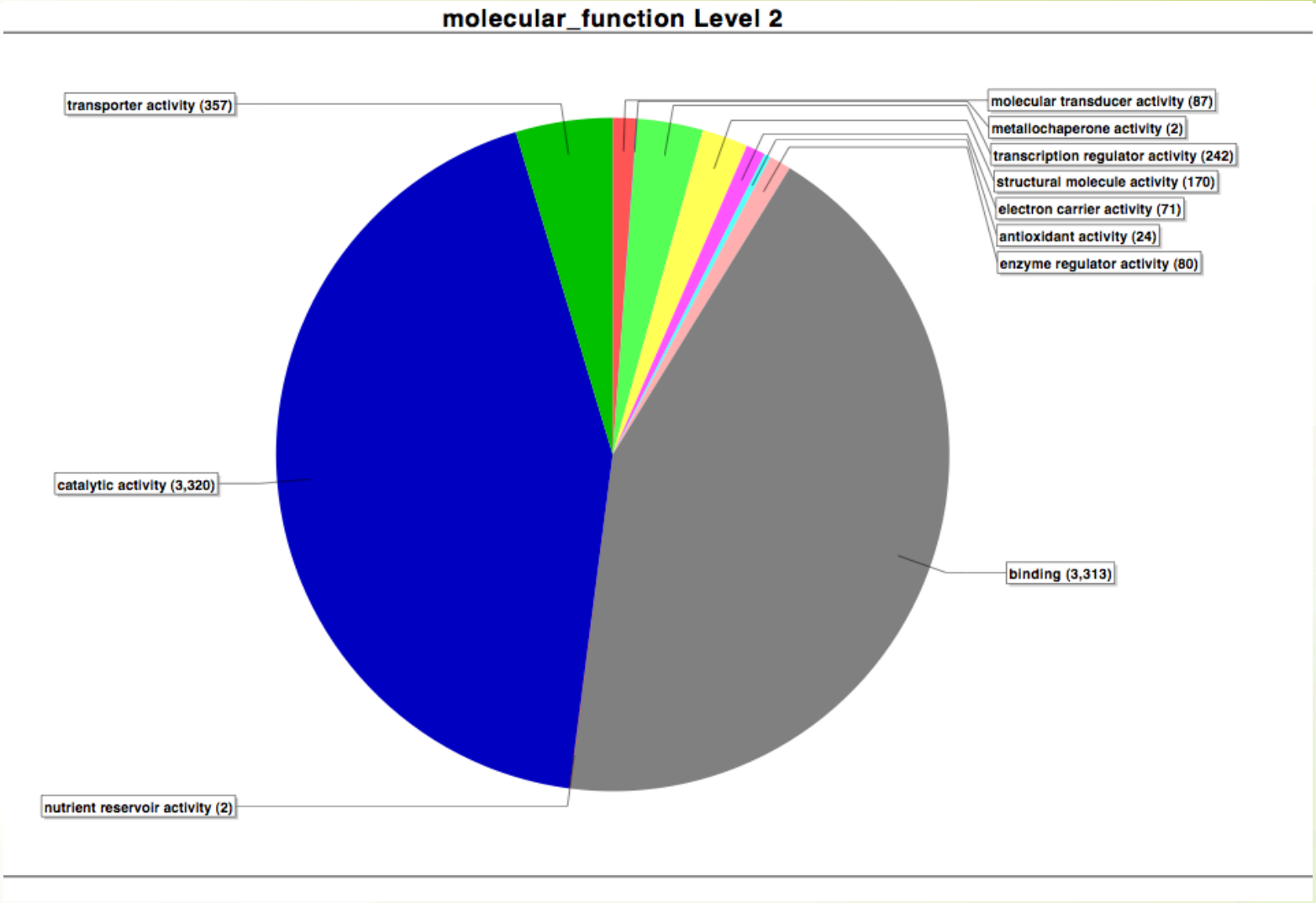
Blast2GO



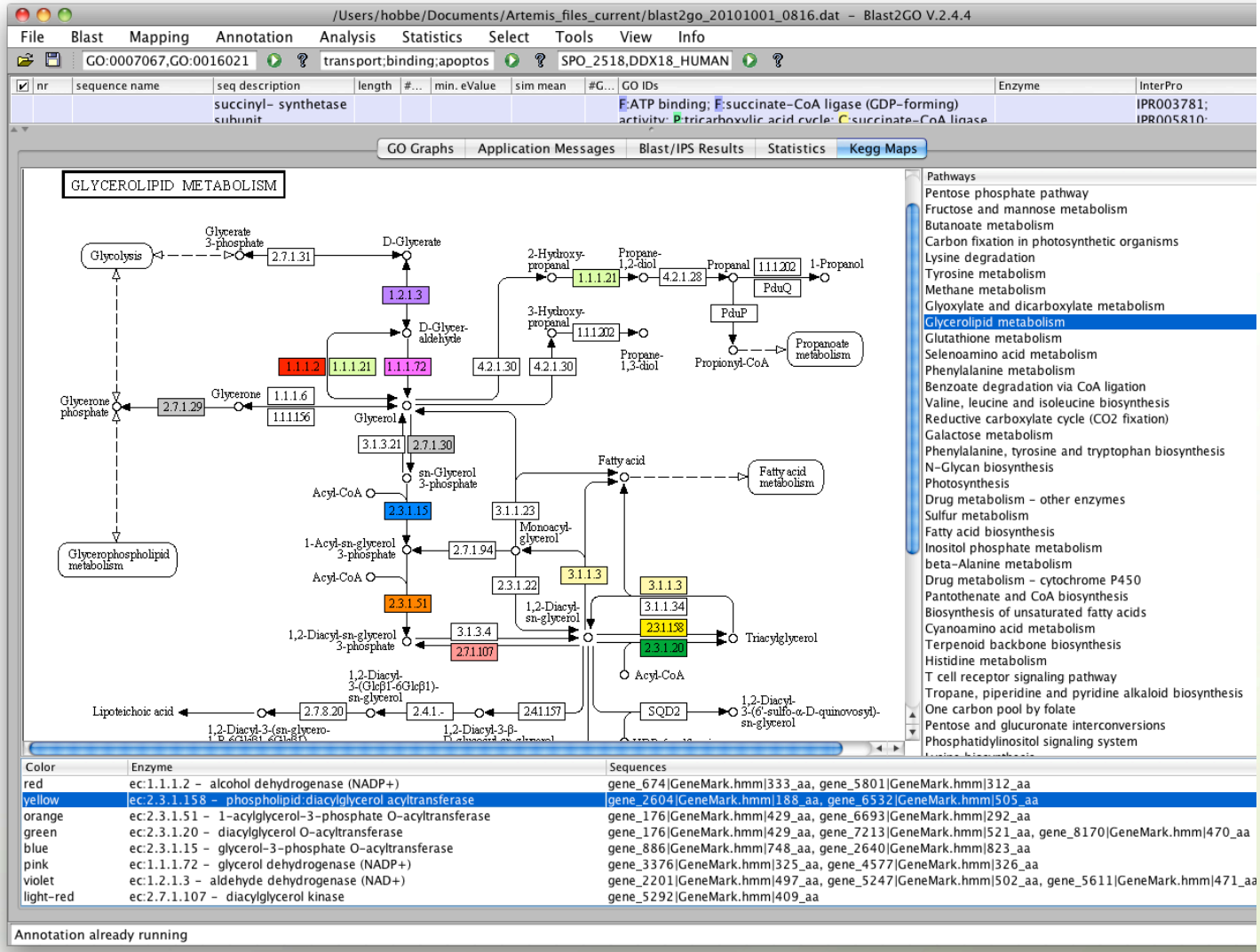
blast2go

Karolinska Institutet · KTH Vetenskap och Konst · Stockholms universitet · Uppsala Universitet

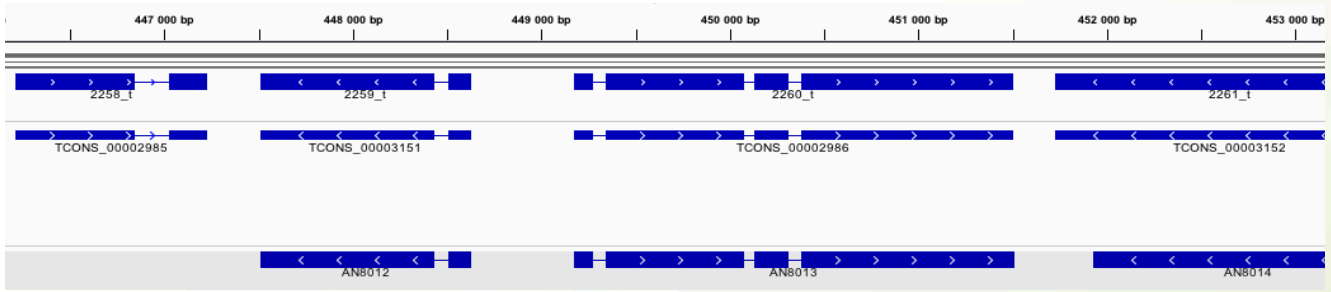
SciLifeLab


molecular_function Level 2
transporter activity (357)
molecular transducer activity (87)
metallochaperone activity (2)
transcription regulator activity (242)
structural molecule activity (170)
electron carrier activity (71)
antioxidant activity (24)
enzyme regulator activity (80)
catalytic activity (3,320)
binding (3,313)
nutrient reservoir activity (2)

KEGG-mapping



Karolinska Institutet · KTH · Stockholms universitet · Uppsala Universitet · SciLifeLab

Liftovers are very useful for orthology determination

- Kraken
- Align the two genomes (Satsuma) and then transfer annotations between aligned regions