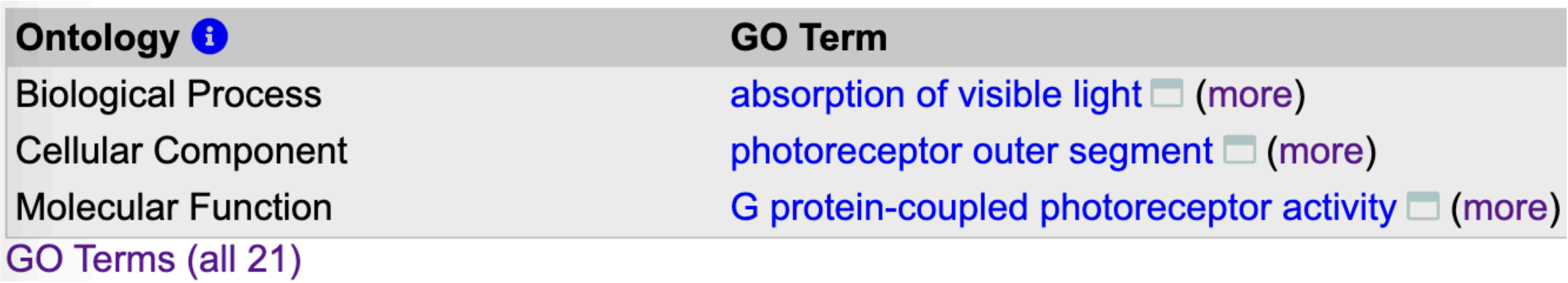
Zebrafish Anatomical Term Enrichment

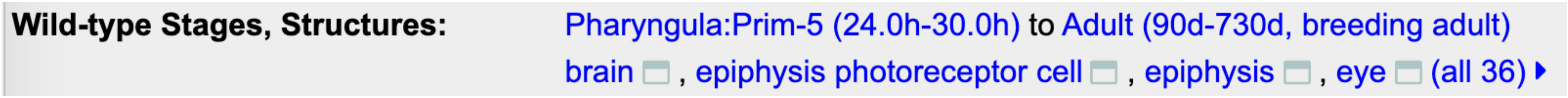
ZFIN annotation

- As well as annotating genes with GO terms, ZFIN associate genes with anatomical terms (based on literature review and in situ images submitted directly)
- For example, rhodopsin (https://zfin.org/ZDB-GENE-990415-271)



Ontology	**GO Term**
Biological Process	absorption of visible light (more)
Cellular Component	photoreceptor outer segment (more)
Molecular Function	G protein-coupled photoreceptor activity (more)

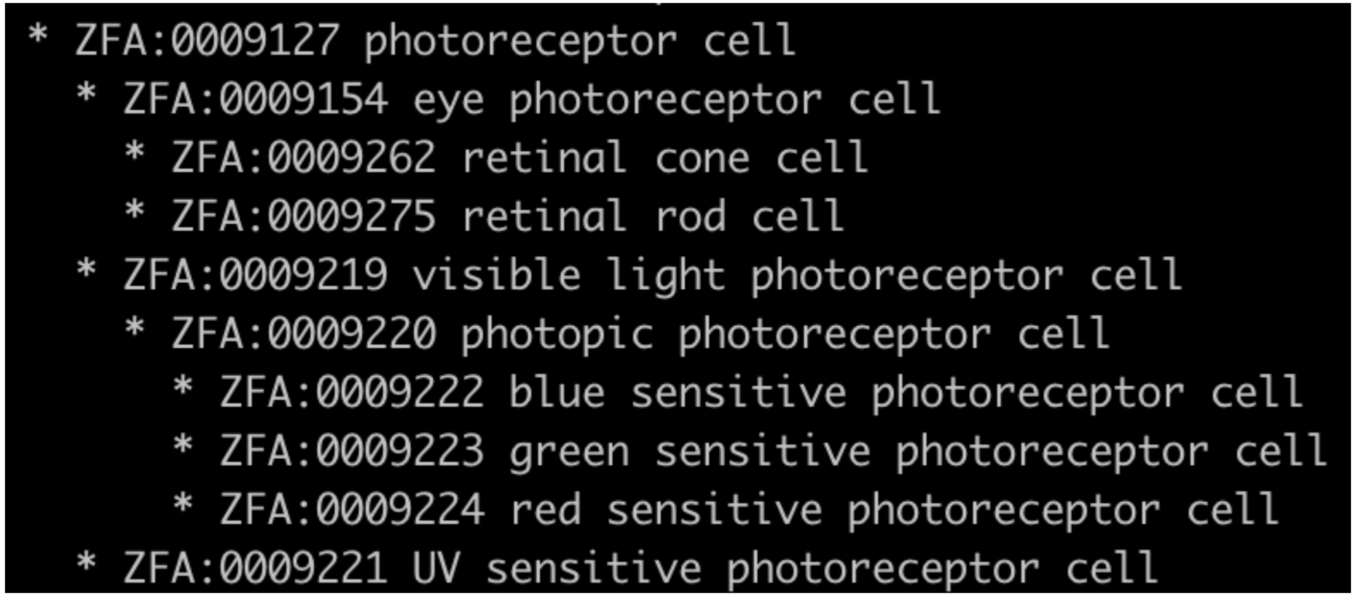
GO Terms (all 21)



Wild-type Stages, Structures:	Pharyngula:Prim-5 (24.0h-30.0h) to Adult (90d-730d, breeding adult) brain, epiphysis photoreceptor cell, epiphysis, eye (all 36) ▸

Anatomical Hierarchy

- The zebrafish anatomy (ZFA) terms form a tree-like hierarchy, where each term can have children and parents

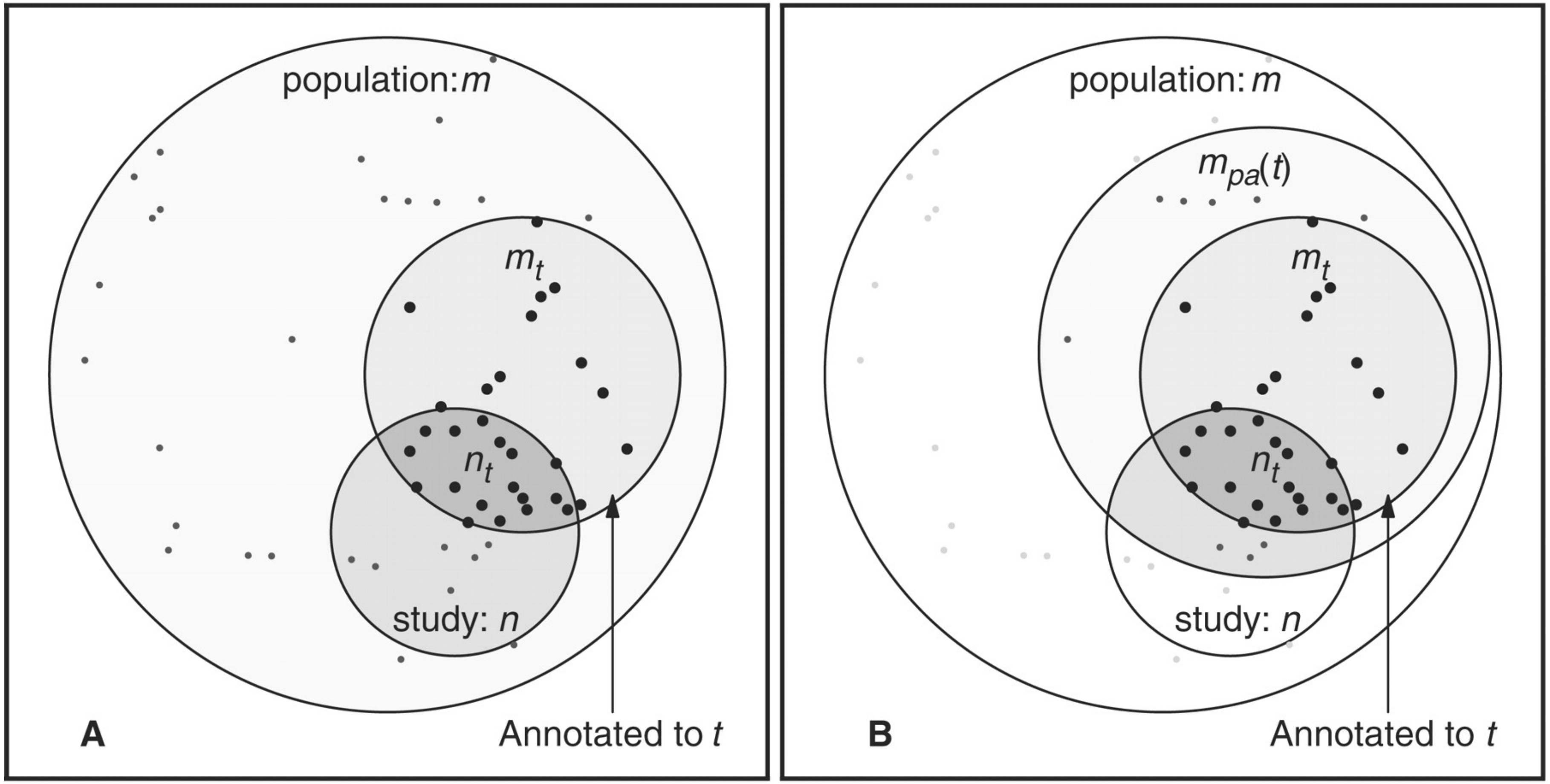


```
* ZFA:0009127 photoreceptor cell
  * ZFA:0009154 eye photoreceptor cell
    * ZFA:0009262 retinal cone cell
    * ZFA:0009275 retinal rod cell
  * ZFA:0009219 visible light photoreceptor cell
    * ZFA:0009220 photopic photoreceptor cell
      * ZFA:0009222 blue sensitive photoreceptor cell
      * ZFA:0009223 green sensitive photoreceptor cell
      * ZFA:0009224 red sensitive photoreceptor cell
  * ZFA:0009221 UV sensitive photoreceptor cell
```

Anatomical Enrichment

- To check for enrichment of terms, can’t just check each term as if they are independent of each other
    - They aren’t, so you’ll get spurious enrichments for some child terms of terms that are significantly enriched
- Instead check for enrichment of a term in the context of the genes annotated to the term’s parents

Anatomical Enrichment



From Grossman et al, 2007

Files

- Need `ensembl.ids` file linking genes to ZFA terms: https://funcgen2019.buschlab.org/downloads/zfa-term-enrichment/ensembl.ids
- And `zfa.obo` file describing all ZFA terms: https://funcgen2019.buschlab.org/downloads/zfa-term-enrichment/zfa.obo
- Both files are also available on ”penelopeprime”

More Files

- Also need a file listing all the Ensembl IDs (the population set):

```
cut -f1 uninf_5dpf_hom_vs_sib.tsv | grep ENS > population.tsv
```

- And a file listing all the significant Ensembl IDs (the study set):

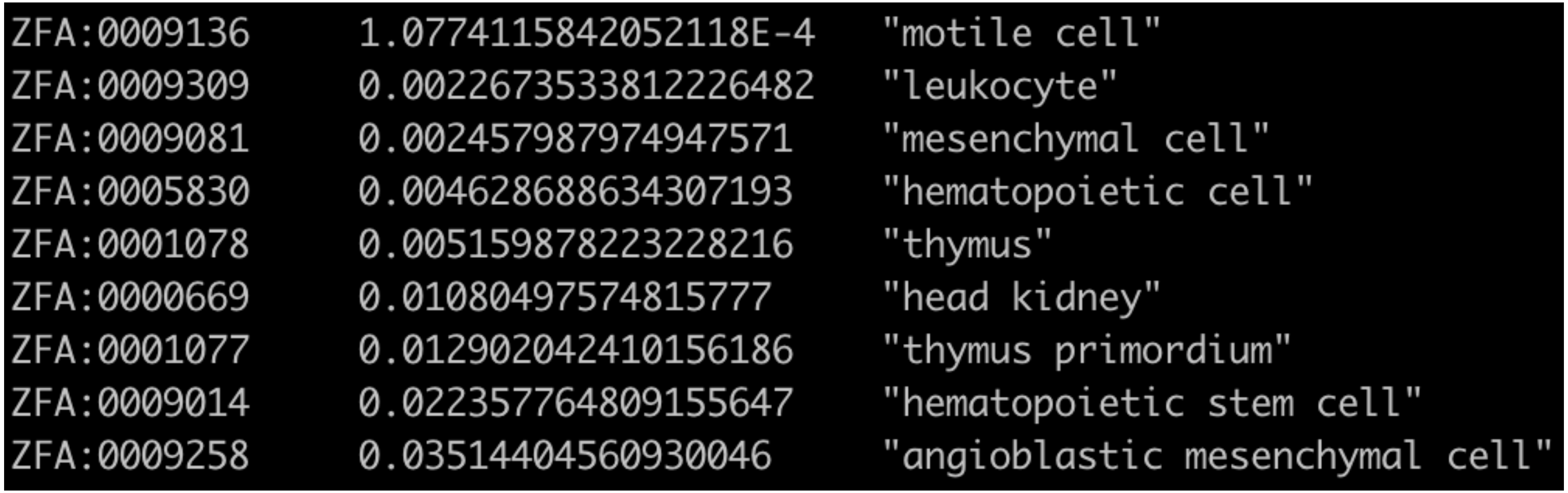
```
cut -f1 uninf_5dpf_hom_vs_sib.sig.tsv | grep ENS > study.tsv
```

Ontologizer

- http://ontologizer.de/ is software for testing enrichment of ontology terms
- The command line version is installed
- `ontologizer -a ensembl.ids -g zfa.obo -c Parent-Child-Union -m Bonferroni -p population.tsv -s study.tsv`
- Will produce a file called `table-study-Parent-Child-Union-Bonferroni.txt` containing all the ZFA terms, including significantly enriched ones

Ontologizer Output

- `cut -f1,11,13 table-study-Parent-Child-Union-Bonferroni.txt | awk '$2 < 0.05'`



```
ZFA:0009136     1.0774115842052118E-4   "motile cell"
ZFA:0009309     0.0022673533812226482   "leukocyte"
ZFA:0009081     0.002457987974947571    "mesenchymal cell"
ZFA:0005830     0.004628688634307193    "hematopoietic cell"
ZFA:0001078     0.005159878223228216    "thymus"
ZFA:0000669     0.01080497574815777     "head kidney"
ZFA:0001077     0.012902042410156186    "thymus primordium"
ZFA:0009014     0.022357764809155647    "hematopoietic stem cell"
ZFA:0009258     0.03514404560930046     "angioblastic mesenchymal cell"
```

Thank You

Any Questions?