

---

# Protein Prediction 2

Exercise  
Team 6

---



---

# Clean Up

---

- ❖ Removed „noise“ from datafiles
- ❖ db.fasta contained duplicated entries
  - ❖ removed duplicates
- ❖ Recreated blast-database



---

# Parsing HPO-Tree

---

- ❖ Tree structure (python)
- ❖ One tree of all hpo-entries in the file
- ❖ Possibility to extract paths from this tree



---

# Creating annotation map

---

- ❖ Creating map / dictionary structure
  - ❖ uniprot-ids / gene-ids  $\longleftrightarrow$  hpo terms



---

# Demo

---

- ❖ Callable by
  - ❖ *gethpo.py <uniprot-id>*
- ❖ Result
  - ❖ List of annotations
  - ❖ Path of each annotation