PROTEIN FUNCTION PREDICTION

- Automated Function Prediction (AFP) tools will play a very important role in medicine and health care in the future.
- However, current tools predict different sets of Gene Ontology (GO) terms for the same input and only few terms are common with the experimentally validated terms.
- Experimentally validated terms are assumed incomplete.
- Consequently, biologists and developers will find difficulty in selecting and testing a tool, respectively.

METHODOLOGY

Metamorphic Testing for AFP tools:

Tyrosinase Canonical Sequence

Source Test Case (T1) = {1,2,3,4,5}
Follow-up Test Case (T2) = {2,4,5,1,3}

Program Under Test (AFP Tool)

GO Terms Set 1 (O1) = {GO:0005737, GO:0042470, GO:0045509, GO:0043231, ...

Check whether the MRs hold for O1 and O2

Tyrosinase Disease Variant

GO Terms Set 2 (O2)

MR: for a disease variant, the generated GO terms should be different (O1 ≠ O2)

Data: Carefully selected proteins and their variants

- Tyrosinase (TYRO_HUMAN)
- Cytokine receptor common subunit gamma (IL2RG_HUMAN)
- Toll-like receptor 4 (TLR4_HUMAN)

Source test case
Follow-up test cases

Canonical sequence

7 Disease Variants
(Albinism)

4 Disease Variants
(severe combined immunodeficiency)

2 Natural Variants
2 Splice Variants

Tools: CAFA top-performers

- FANN
- FANN Multitool (UCL, CONS, and PFP)
- PANDA, CBRG, GORBI
- FunFams
- ProFun
- FunTastic
- FunTastic Suite
- FunTastic-Argos
- FABB
- FAPP
- FunDA

CONCLUSIONS

- Possible to create MRs based on carefully selected protein examples.
- Many tools did not pass all test-cases.
- AFP community need to work together for QA of AFP tools.
- MT seems to be a promising avenue for testing AFP tools.

FUTURE WORK

- Communicate results to authors and post-analysis.
- Explore more specific MRs: "is there a change?" vs. "actual change".
- Develop an expanded test-suite:
  - Explore the feasibility of developing MRs for Cellular Component.
  - Different types of test-cases, different species.
  - Testcases/MRds for HPO term prediction.
- Work with the AFP community to expand the number of tested tools.
- Develop a MT testing framework readily available to users.

ACKNOWLEDGEMENTS

- Function COSI Travel Fellowship.
- PIs and the teams developing the tools: EVEX, Orengo-FunFams, SIFTER, Jones-UCL, Paccanaro Lab, ProFun, PANDA, CBRG, GORBI, FANN-GO, CONS, and PFP.