The GO ontology is widely used for functional annotation of genes and proteins. It describes biological processes (BP), molecular function (MF), and cellular functions (CC) in three distinct hierarchical controlled vocabularies. At the molecular level, functional annotations are often performed by highly conserved parts of proteins, identified by sequence or structure alignments and classified into domains or families (SCOP, CATH, PFAM, TIGRfams, etc.). The InterPro database provides a valuable integrated classification of protein sequences and domains which is linked to nearly all existing other classifications. Interestingly, several InterPro families have been manually annotated with GO terms using expert knowledge and the literature. However, the list of such annotations is incomplete (only 20% of Pfam domains and families possess MF GO functional annotation). Therefore, we developed the GODomainMiner approach to expand the available functional information and data. The GO term prediction approach is a powerful tool for expanding functional annotations of protein domains and families. Based on our ECDomainMiner approach (2), we use the respective associations of protein sequences with GO terms and protein domains to infer direct associations between GO terms and protein domains. Finally, we used our calculated GO Domain associations to devise a systematic, called AutoProf-Announcer, to generate high confidence rules for protein sequence (or structure) annotation.

**Introduction**

**GO-Domain Similarity Matrix**

**GO-DomainMiner**

**AutoProf-Announcer**

**Prediction flowchart**

**GO Domain**

**Prediction Rules**

**Association Rule Generation**

**AutoProf-Announcer**

**Annotation Rules Result**

**Summary of CAFA Challenge Predictions**

**Annotation Examples**

**Conclusion**

Our GODomainMiner approach provides a substantial enrichment of functional annotations at the protein domain level which has been exploited to develop a novel system here called AutoProf-Announcer for protein functional annotation. We used the AutoProf-Announcer to annotate target sequences in CAFA challenge.

**References**


**LITERATURE CITED**

**ACKNOWLEDGMENTS**

This project is funded by the Agence Nationale de la Recherche (grant reference ANR-11-AOME-006-01), Inria and the Lorraine Region. Travel to the meeting was made possible, in part, by a travel award from the NSF.

**Contact**

INRIA Nancy Grand Est, LORIA, Bureau B.133, France. Email: seyed-ziaeddin.albizzi@inria.fr Website: https://members.inria.fr/SAlbizzi/ Phone: +33 (0)7 83 29 89 83