Annotation of HTTP data

Swiss-Prot practices

Swiss Institute of Bioinformatics

Semi-automatic integration of HTP data into Swiss-Prot

- Main HTP data that have been integrated in batch into the UniProtKB concerned PTM detected by Mass-Spectrometry. They are thus not subject to GO annotation .

- Each paper was manually curated, with an emphasis on MS-MS parameters and identification cut-off. Only papers fitting our guidelines were annotated. Sometimes, part of data were ignored because of too low cut-off.

- This manual procedure is currently stopped, as we are working on the use of external proteomics-dedicated databases, such as PRIDE-Q* (for 'Q-rated'), for this kind of annotation.

*http://www.ebi.ac.uk/pride/

Manual integration of HTP data into Swiss-Prot (in test)

- Concerns only the annotation of Arabidopsis thaliana proteins subcellular locations and is not done in batch.
- Occurs during the normal annotation process by looking into the SUBA database*.
- All retrieved PubMed are analyzed. Only papers confirming a predicted or proven subcellular location are kept for UniProtKB.

* http://suba.plantenergy.uwa.edu.au/

Manual integration of HTP data into Swiss-Prot (in test)

- Limitation of HTP data: methods used often introduce contamination.

- Swiss-Prot group decided to not process in batch such results, but to use them as additional proof.
  If any additional evidence is going to the same way, the paper is added to the entry as well as associated subcellular location. But if any proof is in conflict, the annotator has to judge by himself, depending on protein family, predictions, biological context, etc.

- For these reasons, the use of new GO evidence codes (for example NR-IDA vs. R-IDA) may help in recognizing HTP GO annotation, to be aware about potential false positive annotation.

Positive example

ID AB10I_ARATH Reviewed; 271 AA.
AC Q8H1R4; Q9SZC3;
...
DE RecName:
DE Sho
DE Sho
...
DE Flags: Pr
GN Name=AB
GN ORFName
OS Arabidops
...

Ontologies Hide

Keywords

Biological process	Transport
Cellular component	Chloroplast Plastid
Domain	Transit peptide
Ligand	ATP-binding Nucleotide-binding

superfamily. ABCI family.
hain.

[6] "Proteomic stu
Froehlich J.E., W
J. Proteome R
Cited for: SUB

two-dimensional electrophoresis."

HTP data confirmed predictive tools

Sequence a

Featur

Molecule p

Transit

Chain

Biological process	transport Inferred from electronic annotation. Source: UniProtKB-KW
Cellular component	chloroplast envelope Inferred from direct assay. Source: TAIR
Molecular function	ATP binding Inferred from electronic annotation. Source: UniProtKB-KW ATPase activity Inferred from electronic annotation. Source: InterPro

Feature identifier

PRO_0000250664

Negative example

ID CP20C_ARATH Reviewed; 260 AA.
AC P34791; Q8LBZ9; Q9FPH5;

RX PubMed=15821981; DOI=10.1007/s11103-005-0699-3;
RA Giavalisco P., Wilson D., Kreitler T., Lehrach H., Klose J., Gobom J.,
RA Fucini P.;
RT "High heterogeneity within the ribosomal proteins of the Arabidopsis
RT thaliana 80S ribosome.";
RL Plant Mol. Biol. 57:577-591(2005

RX PubMed=18538804; DOI=10.1016/j.phytochem.2008.04.007;
RA Bindschedler L.V., Palmblad M., Cramer R.;
RT "Hydroponic isotope labelling of entire plants (HILEP) for

Some HTP data are in conflict with biological context

J. Biol. Chem. 269:7863-786

Keywords

[7]

Cited for: SUBCELLULAR LO

Domain — Transit peptide

[8] **"Proteomic study of the**

Froehlich J.E., Wilkerson C.
J. Proteome Res. 2:413-425
Cited for: SUBCELLULAR LO

Cellular component

apoplast
Inferred from direct assay. Source: TAIR

chloroplast envelope
Inferred from direct assay. Source: TAIR

chloroplast stroma Ref.1
Inferred from direct assay. Source: TAIR

chloroplast thylakoid membrane
Inferred from direct assay. Source: TAIR

cytosolic ribosome
Inferred from direct assay. Source: TAIR

thylakoid lumen
Inferred from direct assay. Source: TAIR

[9] **"New functions of the thy**

Peltier J.-B., Ytterberg A.J.,
J. Biol. Chem. 279:49367-49
Cited for: SUBCELLULAR LO

Molecule processing

Transit peptide	1 – 78	78	
Chain	79 – 260	182	PRO_0000025476

Conclusion

- At the moment, Swiss-Prot is not annotating in batch HTP papers.
- HTP results are used as decisional tools in complement to other sources.
- To cope with false positives GO annotations, a dedicated evidence code may be used as a 'warning'.