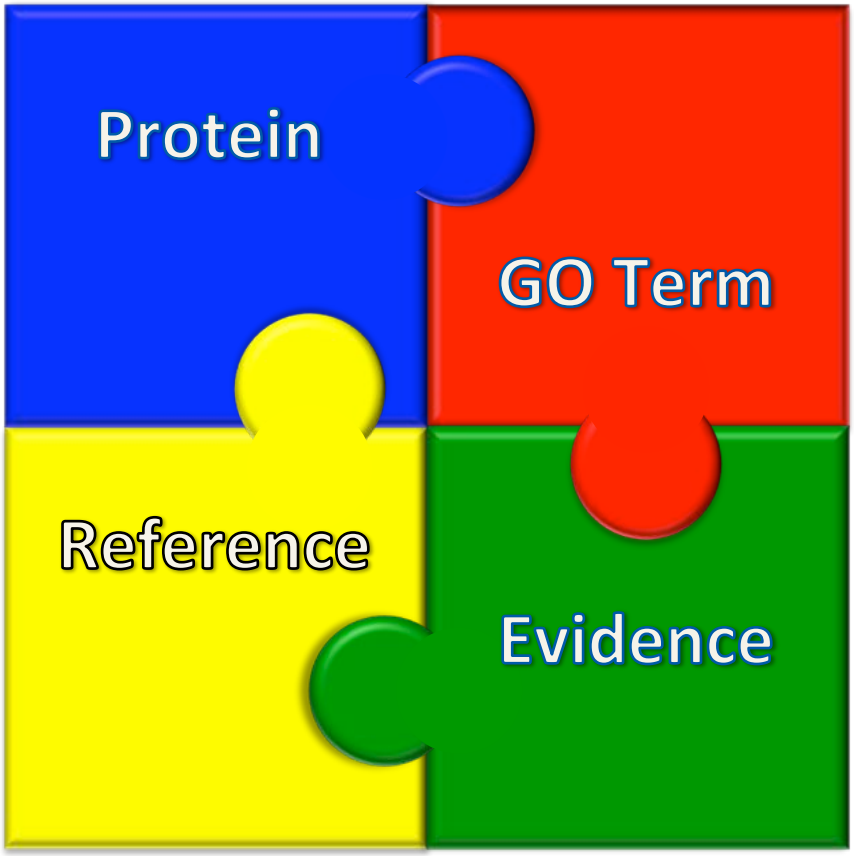
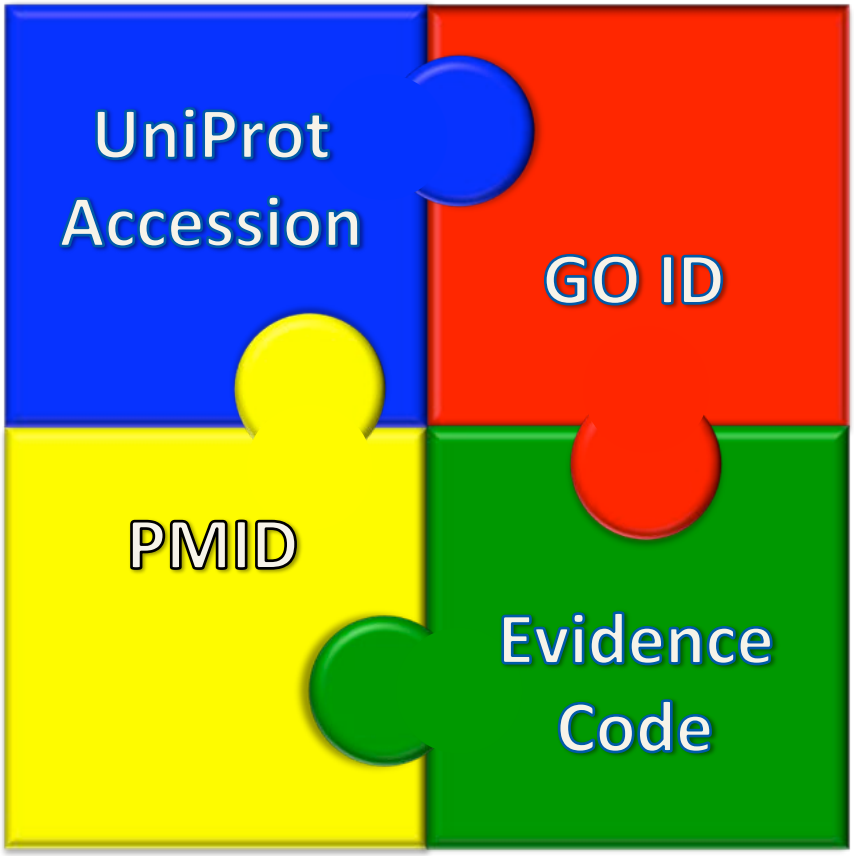
GO Annotation:

Components of an annotation

Components of an Annotation:
What You Need

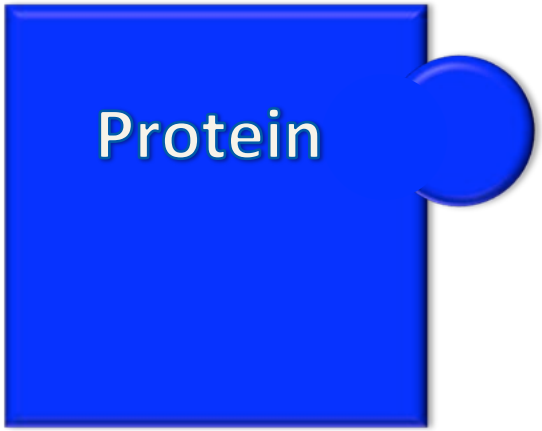


Components of an Annotation:
What You Find



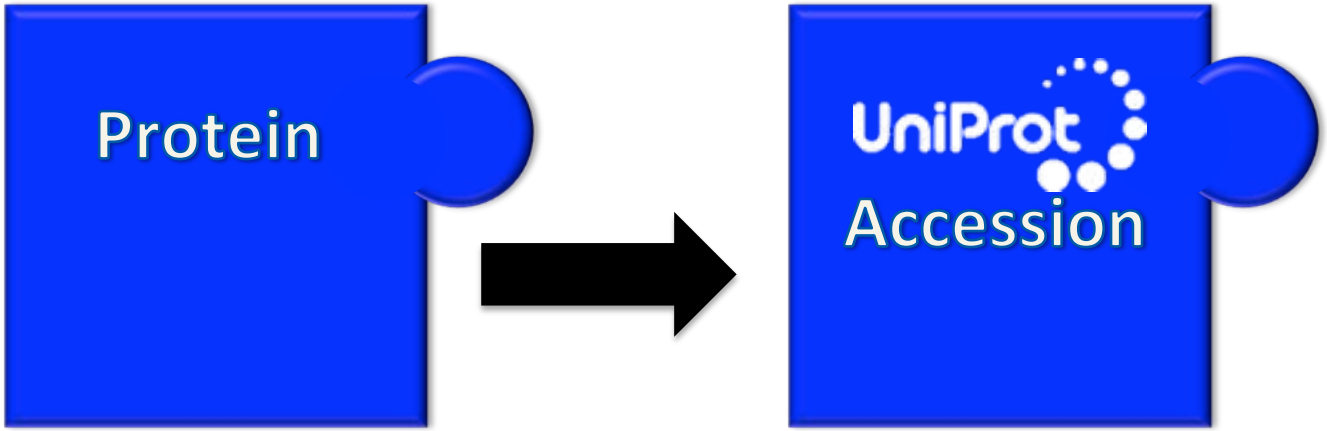
Components of an Annotation:

Protein (Gene Product)

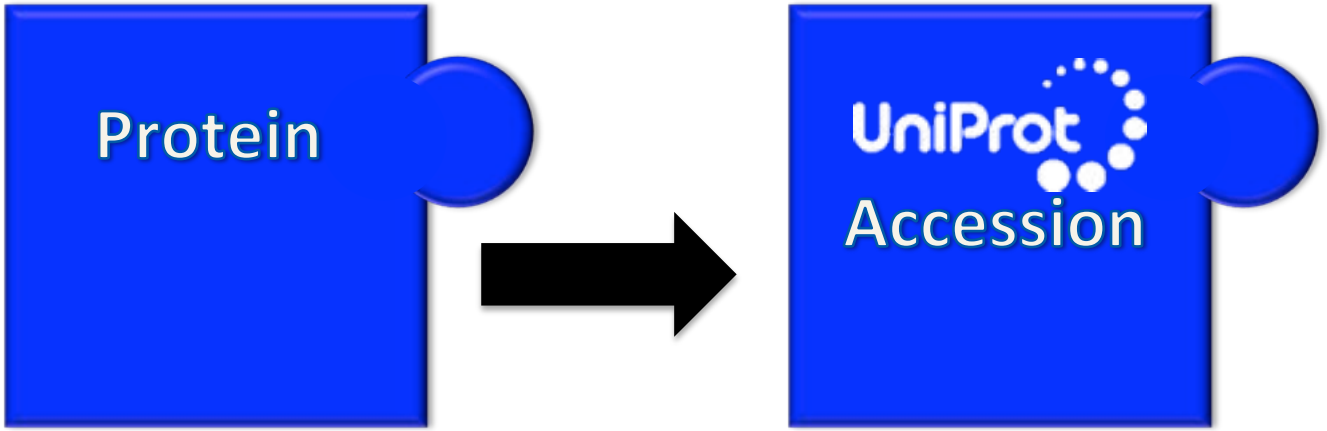


Components of an Annotation:

Protein

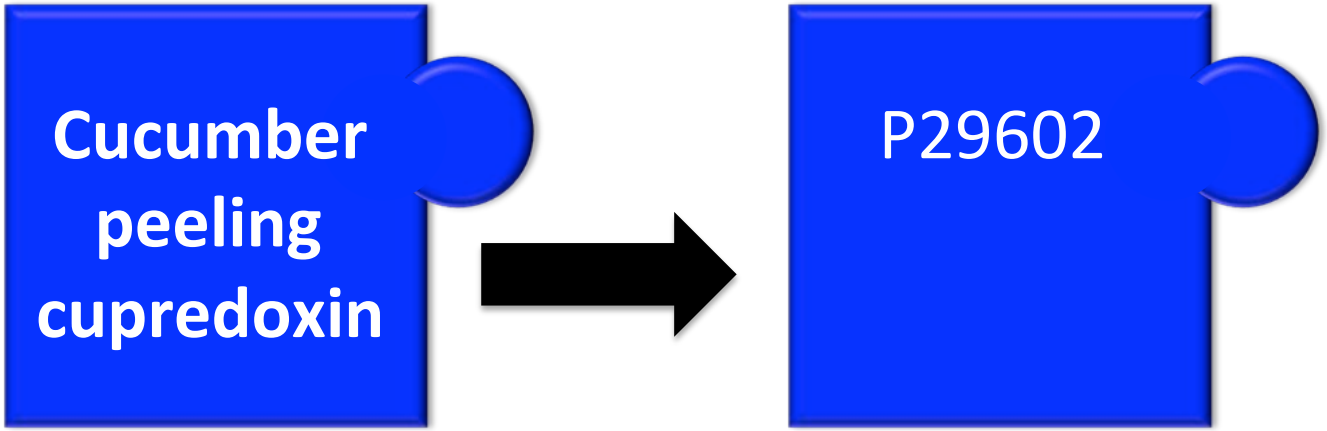


Components of an Annotation:
Protein



- Find the UniProt accession using http://www.uniprot.org/

Components of an Annotation:
Protein



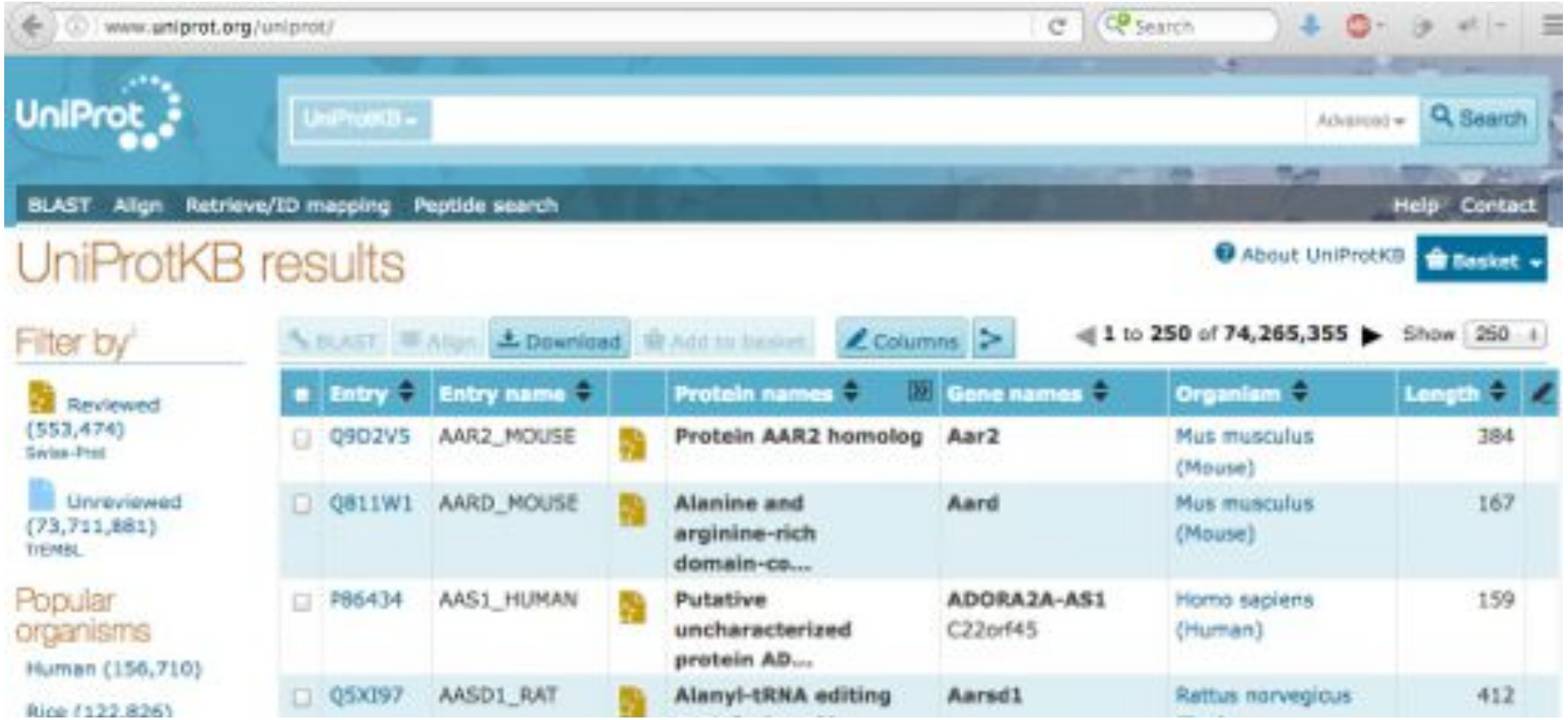
- Find the UniProt accession using http://www.uniprot.org/

- Examples: C4ZZY9, A0A009DWN1

Components of an Annotation:

Protein

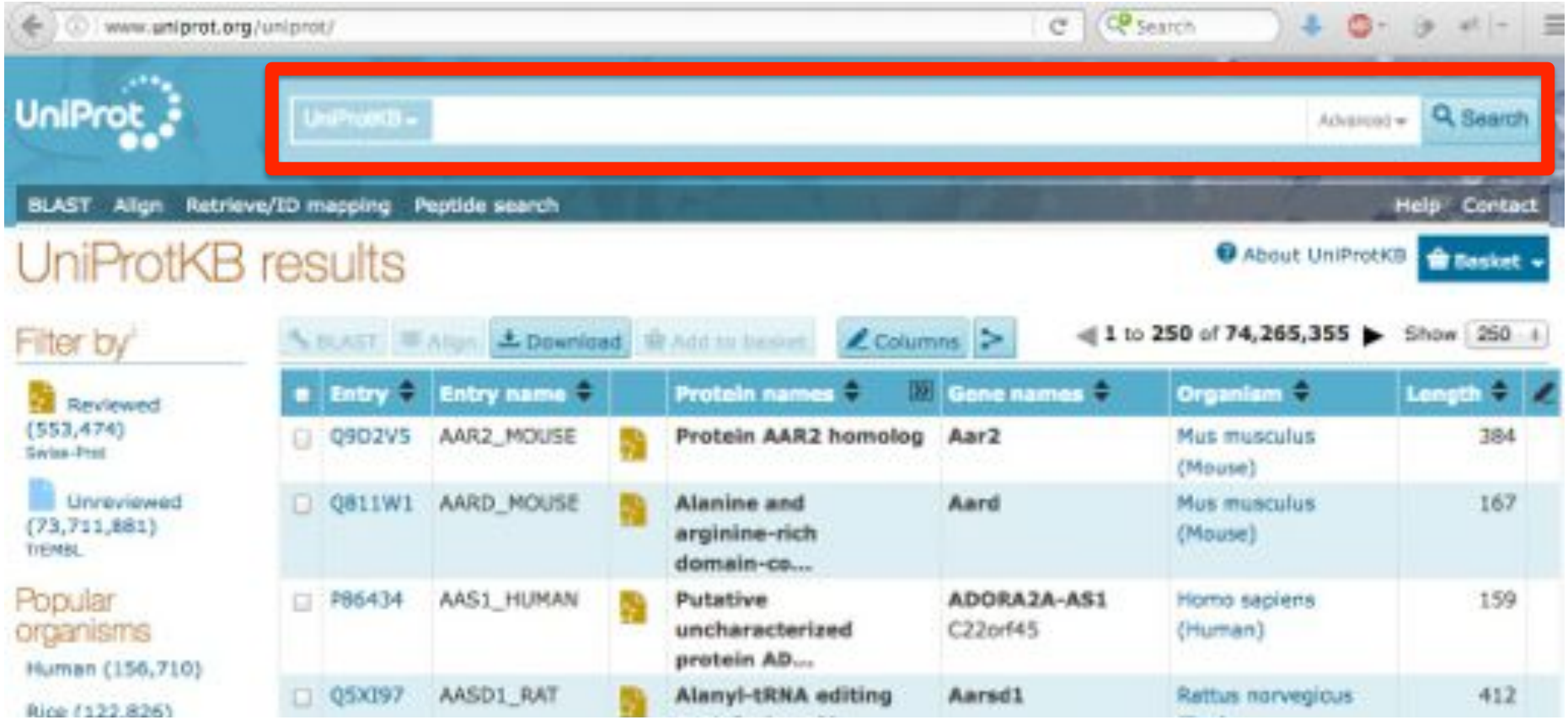
Using UniProt



Components of an Annotation:

Protein

Using UniProt



Components of an Annotation:

Protein

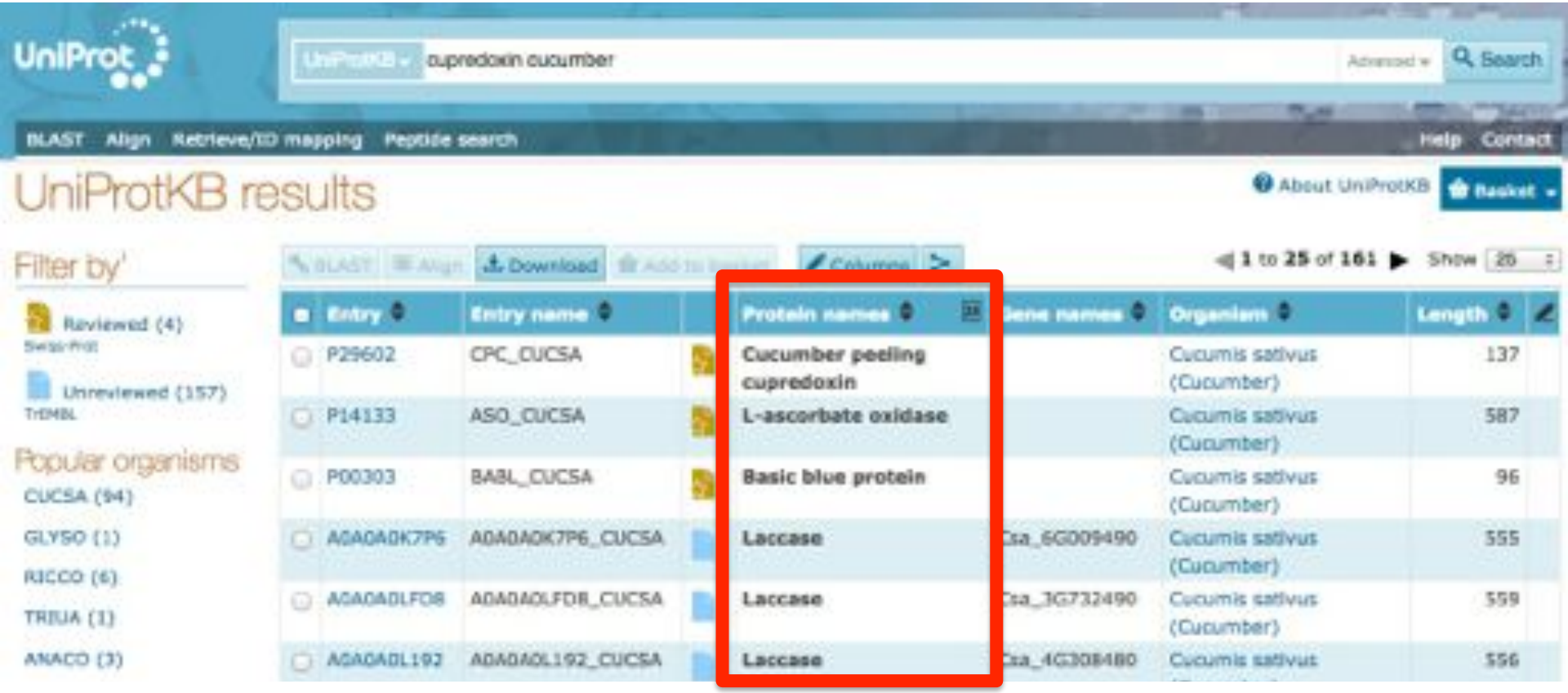
Using UniProt

- Verify the organism (and strain)

Components of an Annotation:

Protein

Using UniProt

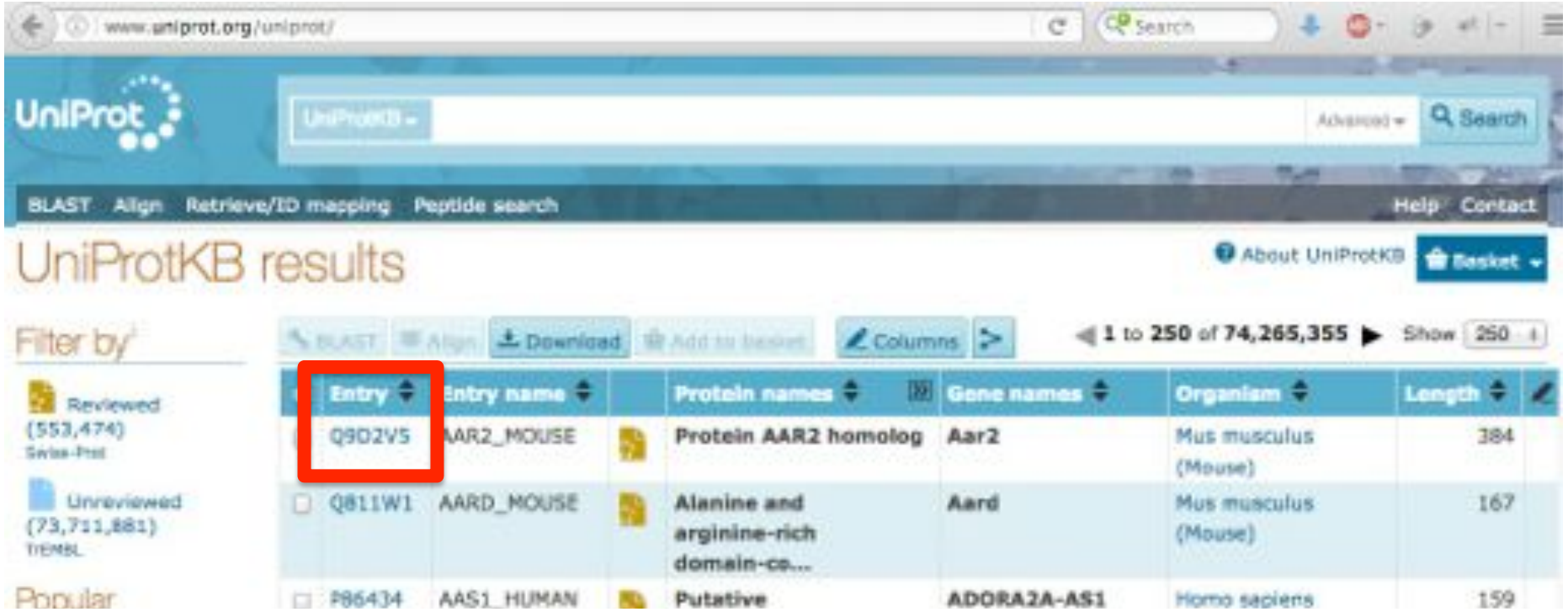


- Verify protein name

Components of an Annotation:

Protein

Using UniProt

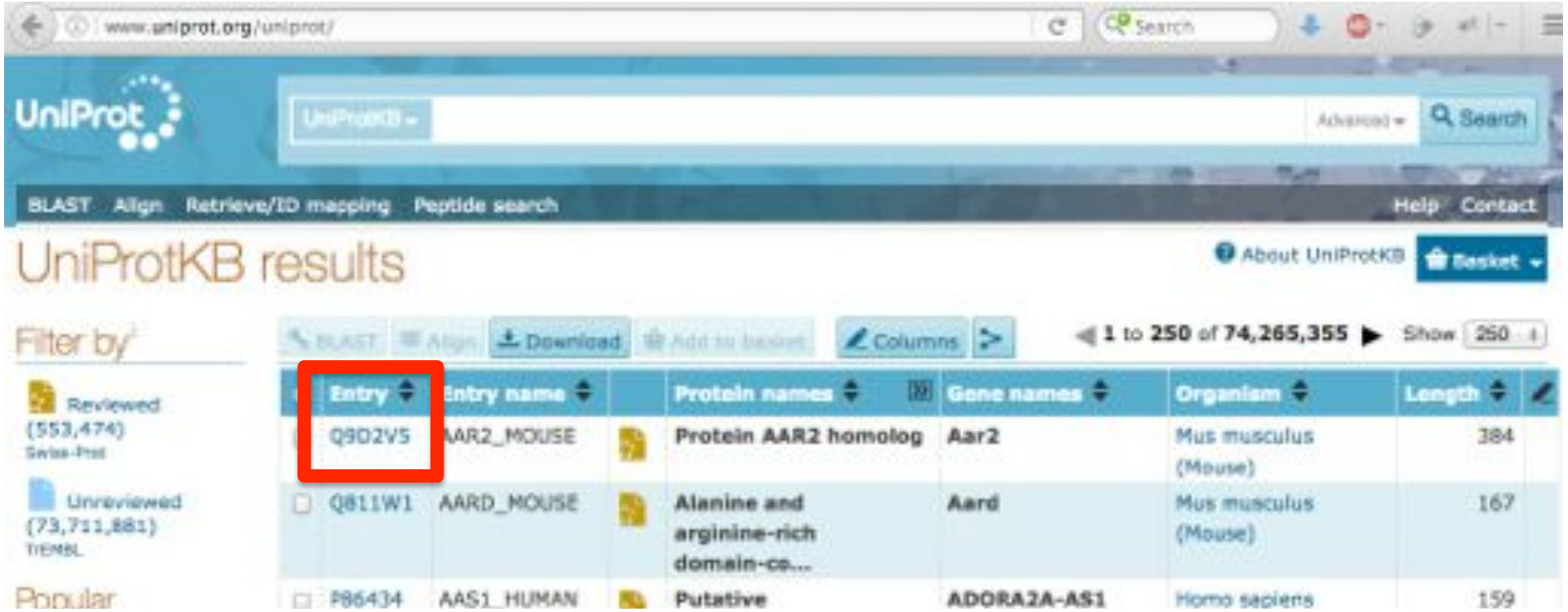


- Copy the accession…

Components of an Annotation:

Protein

Using UniProt



- Copy the accession… *or* ***get more information***

Components of an Annotation:

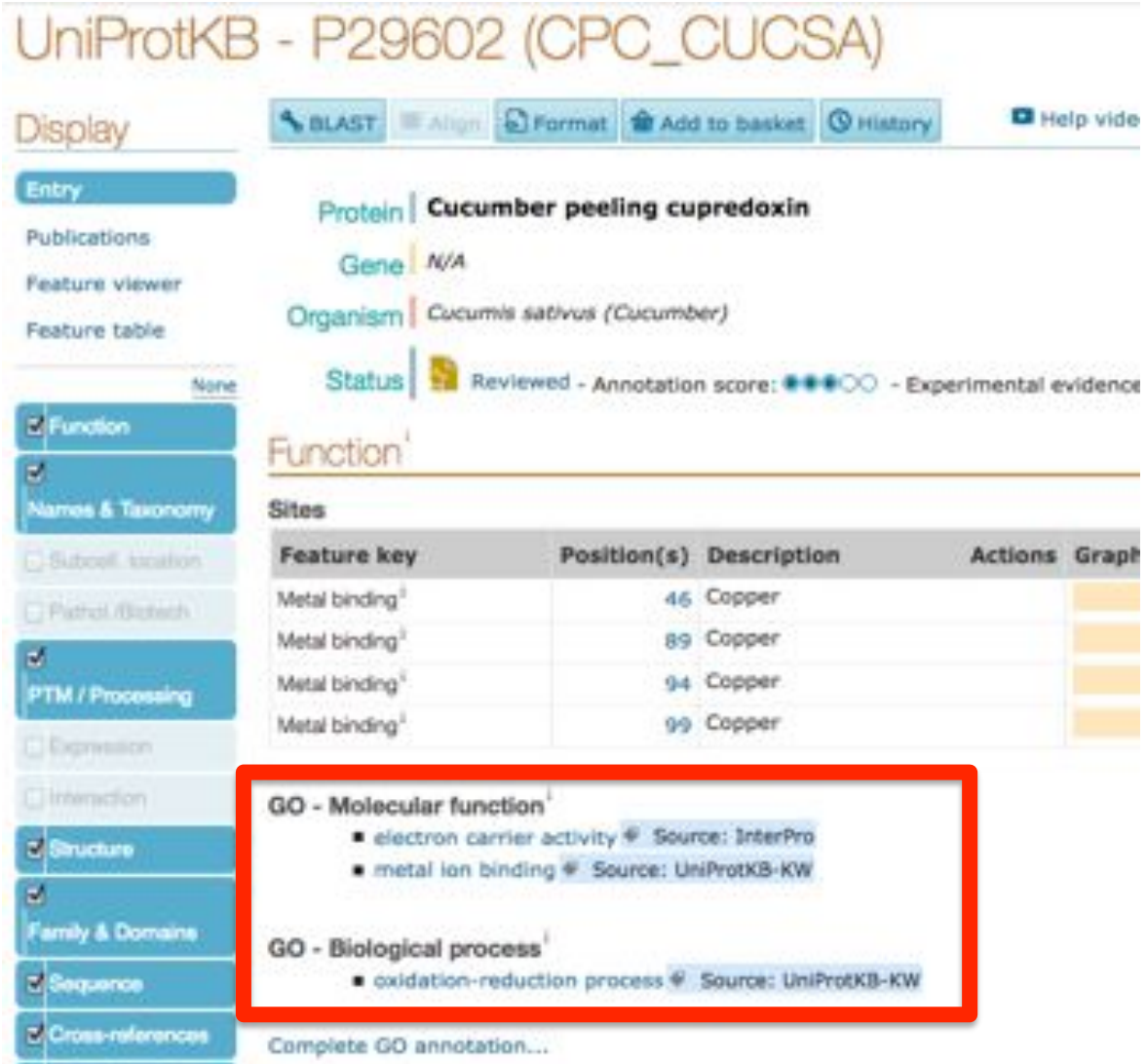
Protein

- Protein record shows more info

Components of an Annotation:

Protein

- Protein record shows more info
- Summary of existing GO annotations



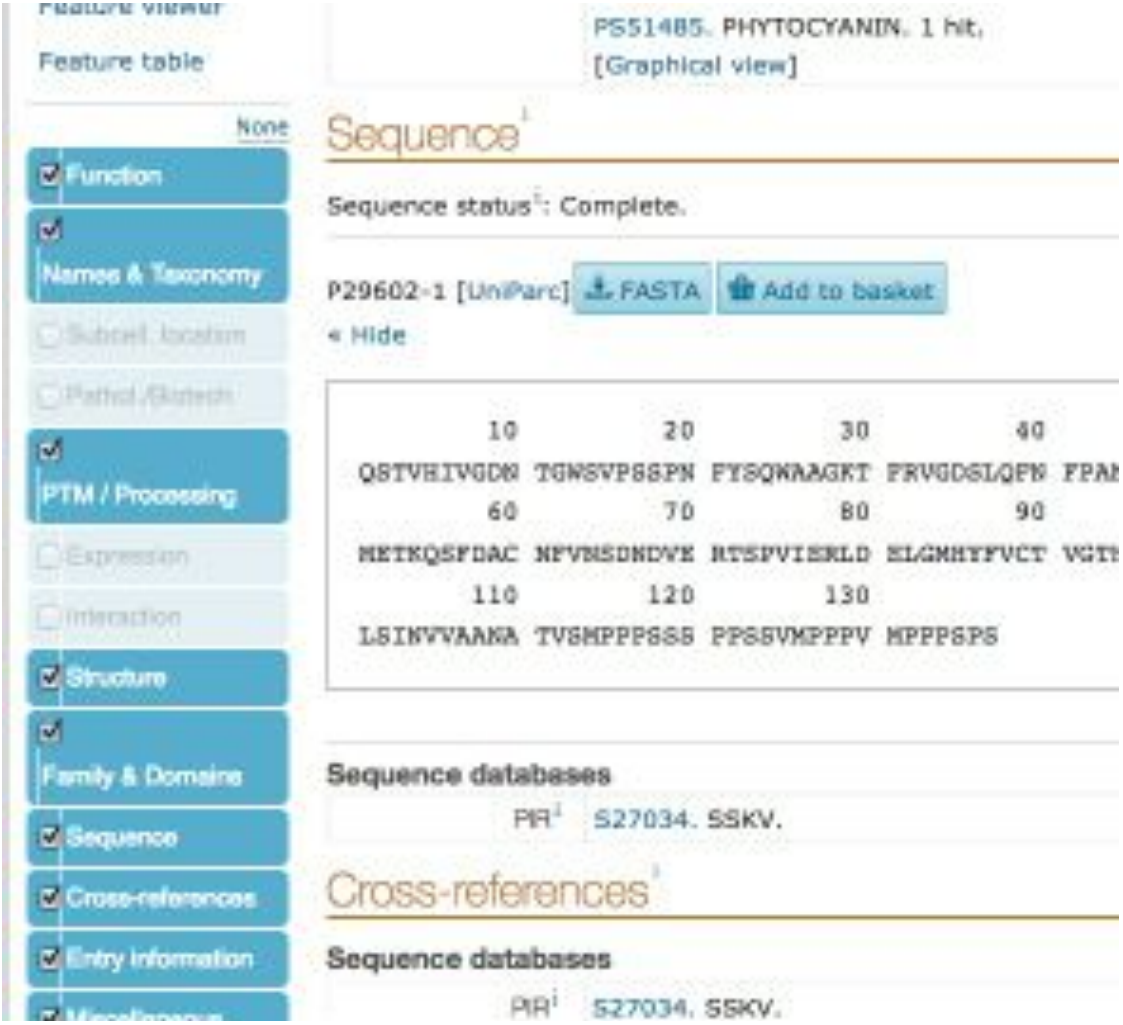
Components of an Annotation:

Protein

- Protein record shows more info
- Summary of existing GO annotations
- Link to QuickGO for complete set of existing annotations

Components of an Annotation:
Protein

- Protein record shows more info
- Sequence & length
- Cross References
- Etc.



Components of an Annotation:
Protein

- Look in the paper for any accession/ID
  - UniProt accession
    - E2RF02
  - NCBI Gene ID, RefSeq, etc.
    - 478211
    - XP_853095.1
    - NP_001153791.1

Components of an Annotation:
Protein

- E2RF02

- You MUST use the UniProt accession when adding the annotation to CACAO

- 478211
- XP_853095.1
- NP_001153791.1

- Papers may use identifiers other than UniProt accessions
- Use these to find the UniProt accession

Components of an Annotation:

Protein

UniProt Warnings

- Sometimes UniProt has multiple entries for the same protein
  - Gold star = SwissProt = reviewed
  - Blank star = TrEMBL = computational entry
- **Ask for help**

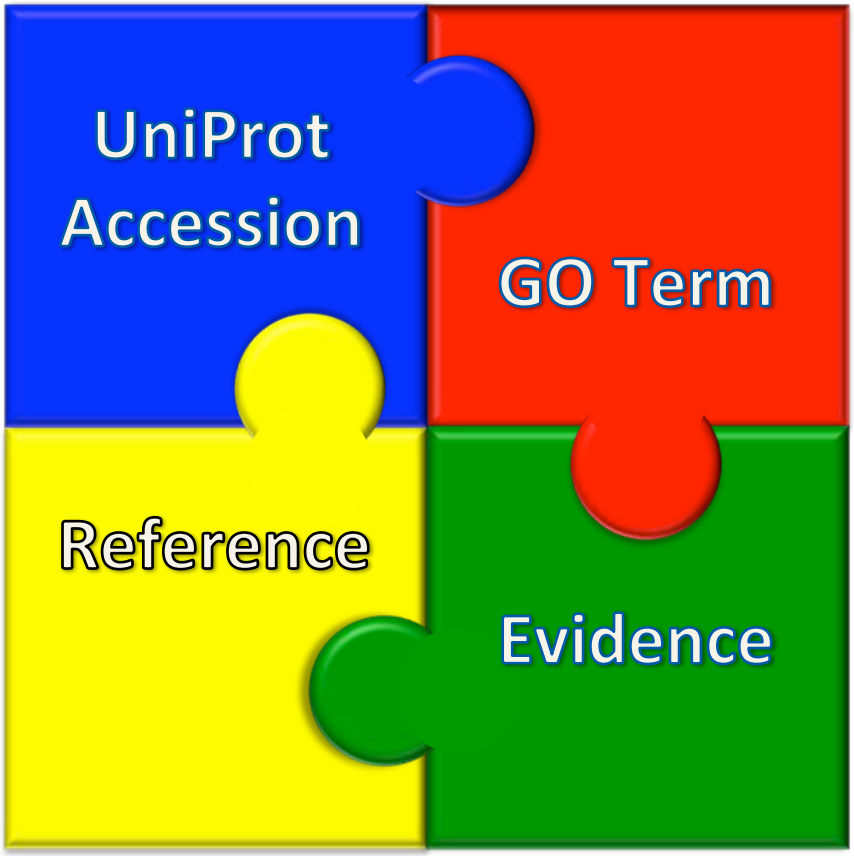
Components of an Annotation:

Protein

UniProt Warnings

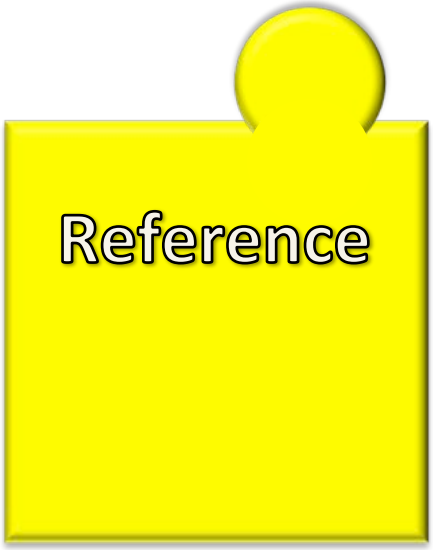
- Right species/strain
- Not a fragment
- Sometimes the protein you want is not in UniProt
  - May want to find another paper/protein
- **Ask for help**
  - OK to email the UniProt help desk
  - Check your reasoning with us!

Components of an Annotation:
What We've Covered



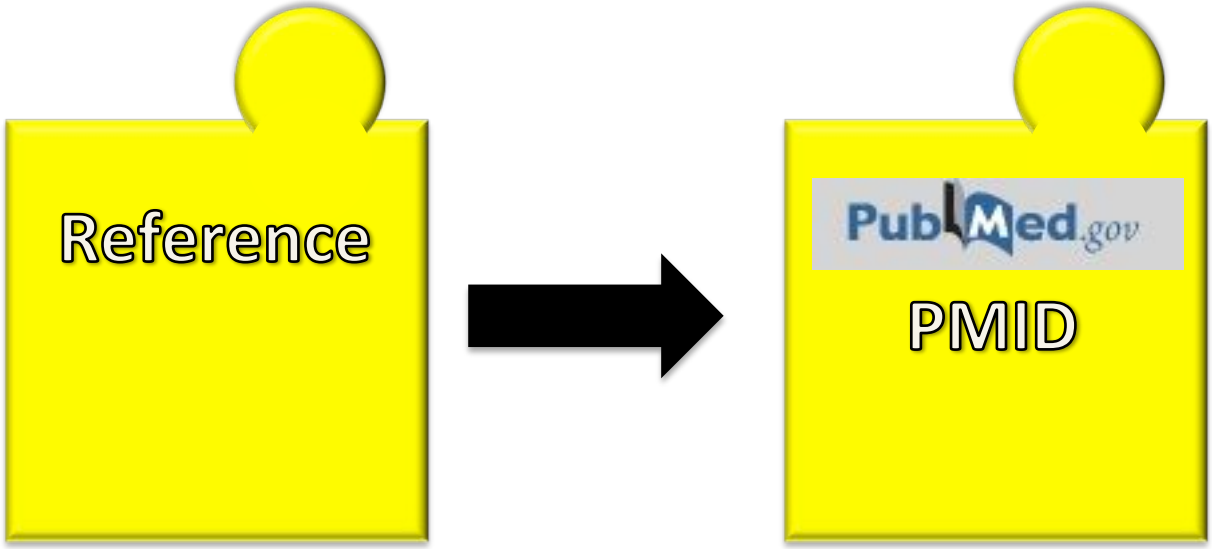
Components of an Annotation:

Reference



Components of an Annotation:

Reference



- Find the PMID using **pubmed.gov** (https://www.ncbi.nlm.nih.gov/pubmed/)

- Examples: **PMID:**22110029, **PMID:**20473289

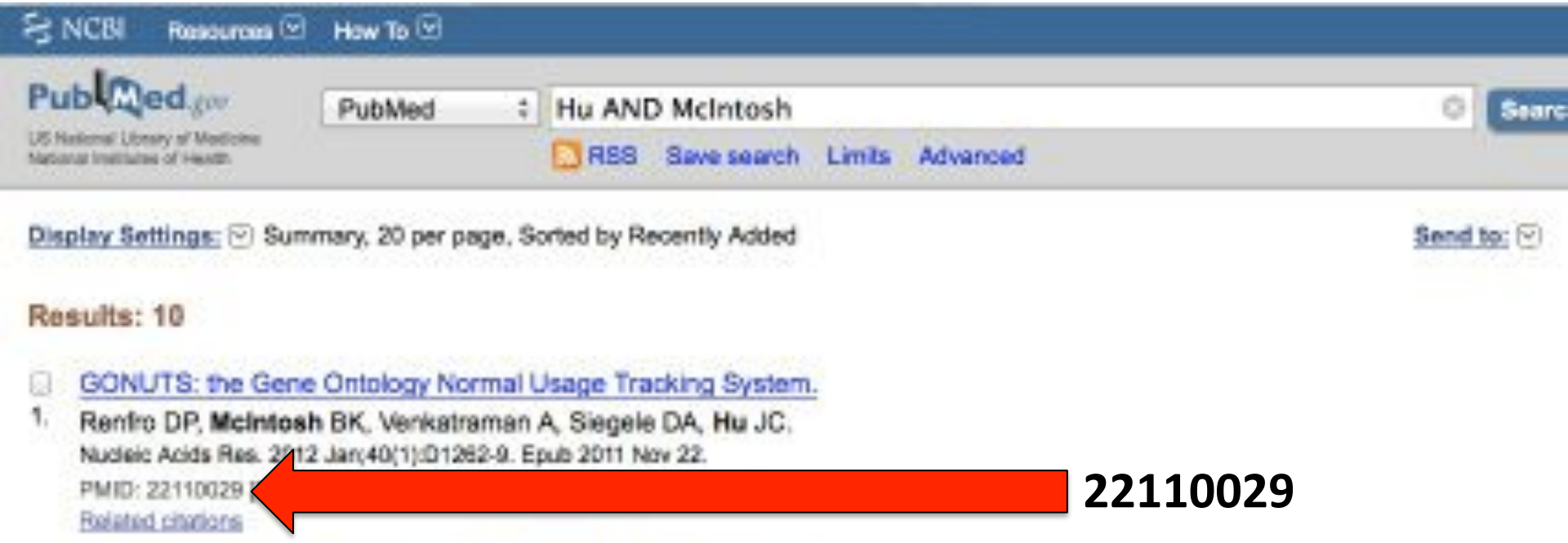
Components of an Annotation:
Reference

- Need a scientific paper with experimental data

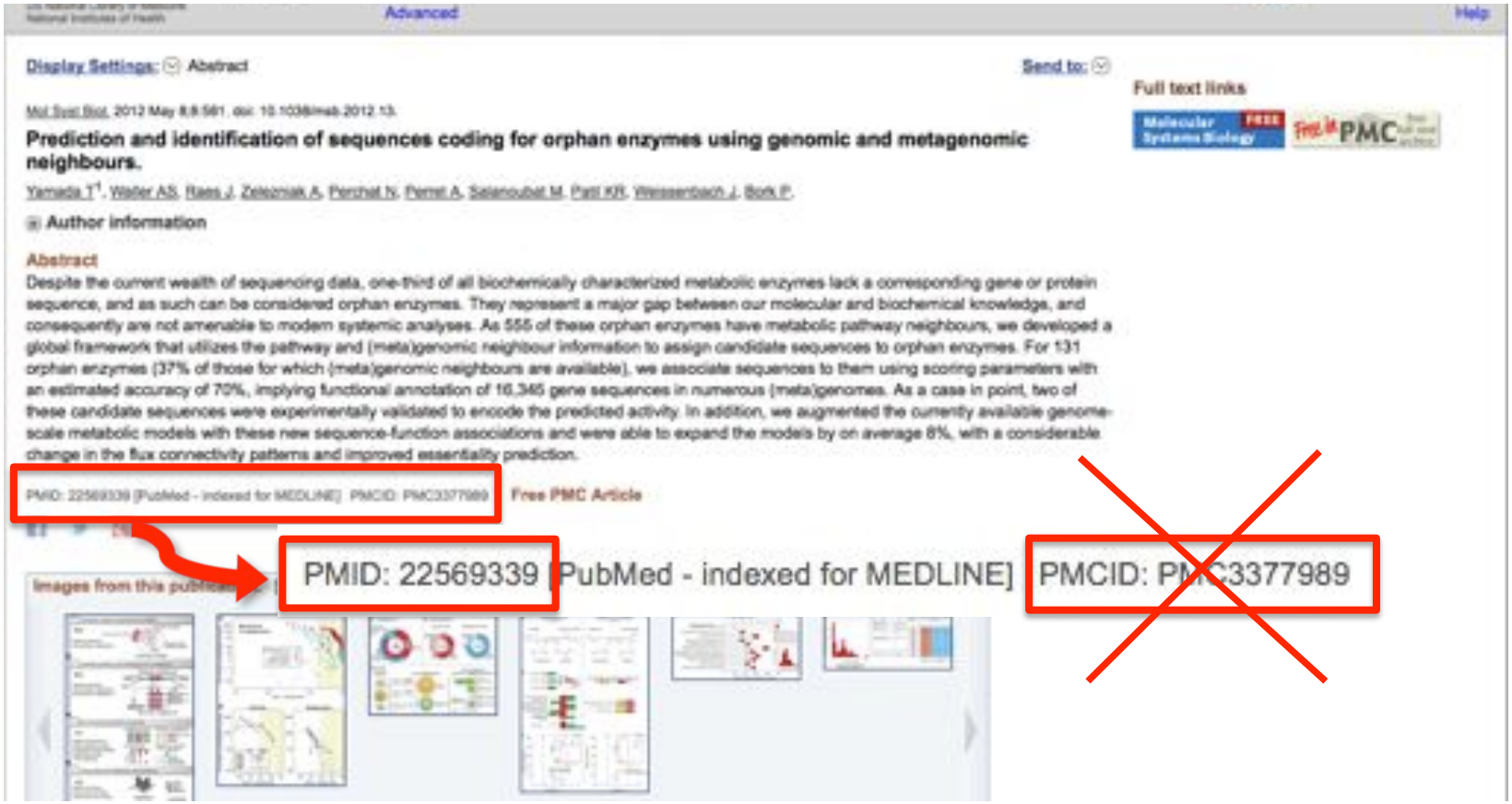
- No review articles, no books, no textbooks, no Wikipedia articles, no class notes…
  - BUT it is GOOD to start with those!
- DON'T start with the first paper you see from a random PubMed search

Components of an Annotation:
Reference

- Need a scientific paper with experimental data
  - PubMed review?
  - Refer to the paper through the PMID number
    - *Not* the full citation



Components of an Annotation:
Reference



Components of an Annotation:
Reference

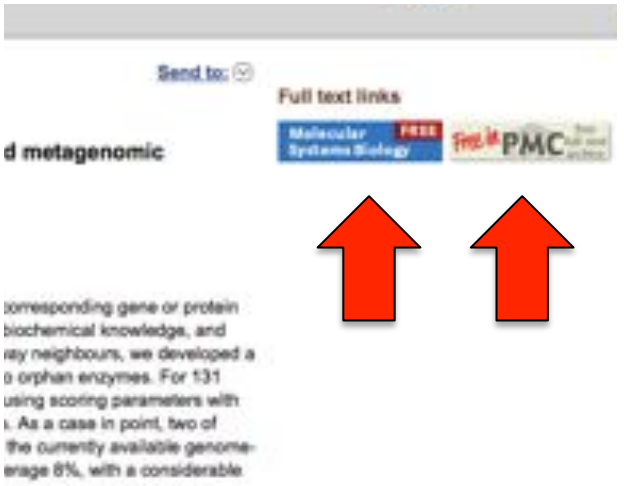
- The abstract is not enough for an annotation
  - But, may be enough to reject a paper!!!

Components of an Annotation:
Reference

- PubMed has links to full paper

Components of an Annotation:
Reference

- PubMed has links to full papers
  - Links vary between papers
- Some papers are open access
  - Pubmed Central
  - Journal sites
  - TAMU students:
    http://library.tamu.edu/
- Others are pay only
  - Don't pay real $$!
    - Your library may have subscriptions
    - Pick a different paper
    - Email the author and ask for a pdf
      - Send us a copy



Components of an Annotation:

Reference

Warnings

- Good science ≠ good for annotation

Second Extracellular Loop of Human Glucagon-like Peptide-1 Receptor (GLP-1R) Differentially Regulates Orthosteric but Not Allosteric Agonist Binding and Function[*] [S]

Received for publication, September 30, 2011, and in revised form, November 29, 2011 Published, JBC Papers in Press, December 6, 2011, DOI 10.1074/jbc.M111.309369

Cassandra Koole[‡], **Denise Wootten**[‡], **John Simms**[‡], **Emilia E. Savage**[‡], **Laurence J. Miller**[§], **Arthur Christopoulos**[‡1], **and Patrick M. Sexton**[‡2]

From the [‡]*Drug Discovery Biology, Monash Institute of Pharmaceutical Sciences and Department of Pharmacology, Monash University, Parkville, Victoria 3052, Australia and the* [§]*Department of Molecular Pharmacology and Experimental Therapeutics, Mayo Clinic, Scottsdale, Arizona 85259*

Background: The ECL2 of the GLP-1R is critical for GLP-1 peptide-mediated selective signaling.
Results: Mutation of most ECL2 residues to alanine results in changes in binding and/or efficacy of oxyntomodulin and exendin-4 but not allosteric agonists.
Conclusion: ECL2 of the GLP-1R has ligand-specific as well as general effects on peptide agonist-mediated receptor activation.
Significance: This work provides insight into control of family B GPCR activation transition.

Components of an Annotation:

Reference

Warnings

- Good science ≠ good for annotation

Cell Stem Cell
Short Article

Cell PRESS

Vitamin C Enhances the Generation of Mouse and Human Induced Pluripotent Stem Cells

Miguel Angel Esteban,[1,6] Tao Wang,[1,6] Baoming Qin,[1,6] Jiayin Yang,[1] Dajiang Qin,[1] Jinglei Cai,[1] Wen Li,[1] Zhihui Weng,[1] Jiekai Chen,[1] Su Ni,[1] Keshi Chen,[1] Yuan Li,[1] Xiaopeng Liu,[1] Jianyong Xu,[1] Shiqiang Zhang,[1] Feng Li,[1] Wenzhi He,[1] Krystyna Labuda,[2] Yancheng Song,[3] Anja Peterbauer,[4] Susanne Wolbank,[2] Heinz Redl,[2] Mei Zhong,[5] Daozhang Cai,[3] Lingwen Zeng,[1] and Duanqing Pei[1,*]

[1]Stem Cell and Cancer Biology Group, Key Laboratory of Regenerative Biology, South China Institute for Stem Cell Biology and Regenerative Medicine, Guangzhou Institutes of Biomedicine and Health, Chinese Academy of Sciences, Guangzhou 510663, China

[2]Ludwig Boltzmann Institute for Clinical and Experimental Traumatology, Austrian Cluster for Tissue Regeneration, Vienna 1200, Austria

Components of an Annotation:

Reference

Warnings

- Good science ≠ good for annotation

Robust design and optimization of retroaldol enzymes

Eric A. Althoff,[1,2] Ling Wang,[1] Lin Jiang,[1,3] Lars Giger,[4] Jonathan K. Lassila,[5] Zhizhi Wang,[1] Matthew Smith,[1] Sanjay Hari,[1] Peter Kast,[4] Daniel Herschlag,[5] Donald Hilvert,[4] and David Baker[1]*

[1]Department of Biochemistry, University of Washington and HHMI, Seattle, Washington 98195
[2]Arzeda Corp., Seattle, Washington 98102
[3]Department of Biological Chemistry, UCLA, Los Angeles, California 90095
[4]Laboratory of Organic Chemistry, ETH Zurich, 8093 Zurich, Switzerland
[5]Department of Biochemistry, Stanford University, Stanford, California 94305

Components of an Annotation:

Reference

Warnings

- Good science ≠ good for annotation

10624 • The Journal of Neuroscience, August 11, 2010 • 30(32):10624–10638

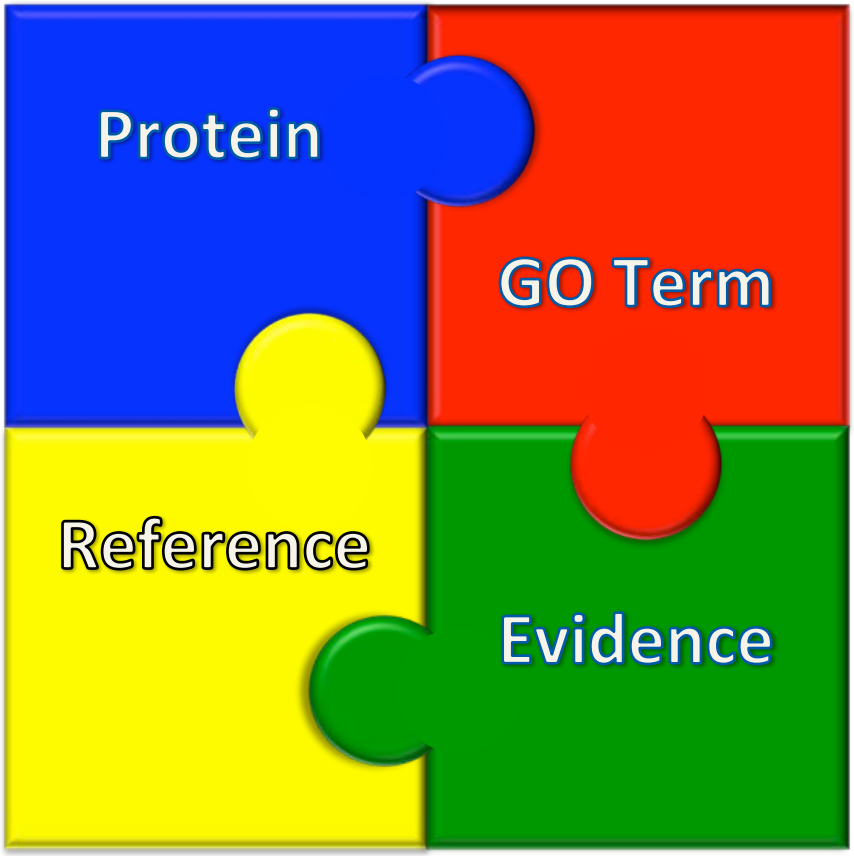
Neurobiology of Disease

Excess Phosphoinositide 3-Kinase Subunit Synthesis and Activity as a Novel Therapeutic Target in Fragile X Syndrome

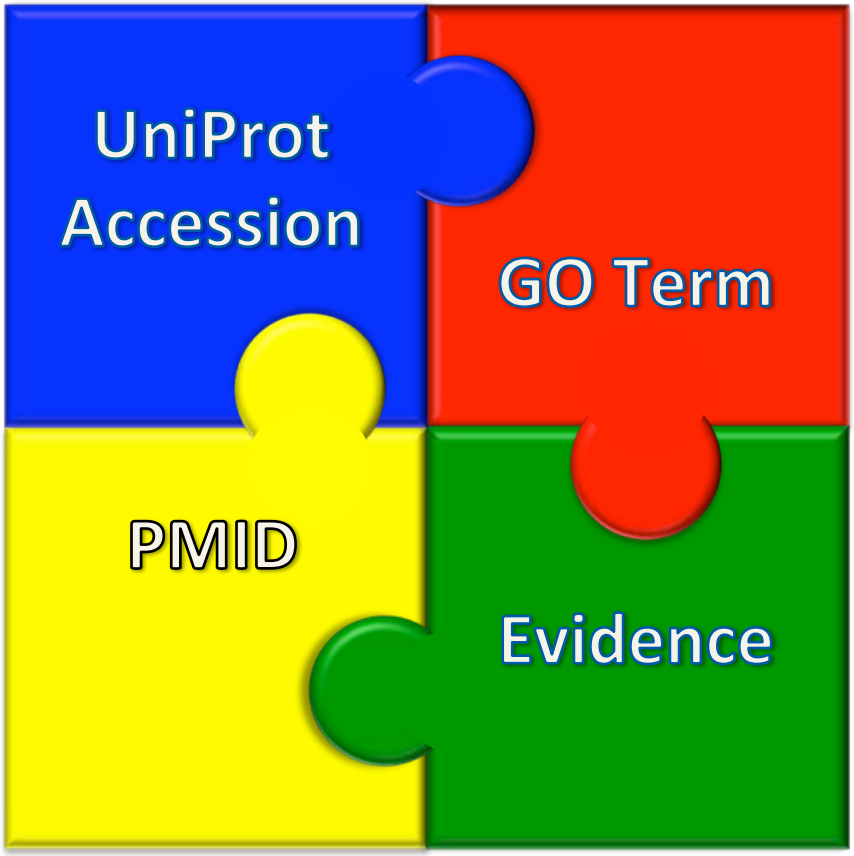
Christina Gross,[1] Mika Nakamoto,[2]* Xiaodi Yao,[1]* Chi-Bun Chan,[3] So Y. Yim,[1] Keqiang Ye,[3] Stephen T. Warren,[2,4,5] and Gary J. Bassell[1,6]

Departments of [1]Cell Biology, [2]Human Genetics, [3]Pathology and Laboratory Medicine, [4]Biochemistry, [5]Pediatrics, and [6]Neurology, Emory University School of Medicine, Atlanta, Georgia 30322

Components of an Annotation:
What We've Covered



Components of an Annotation:
What We've Covered

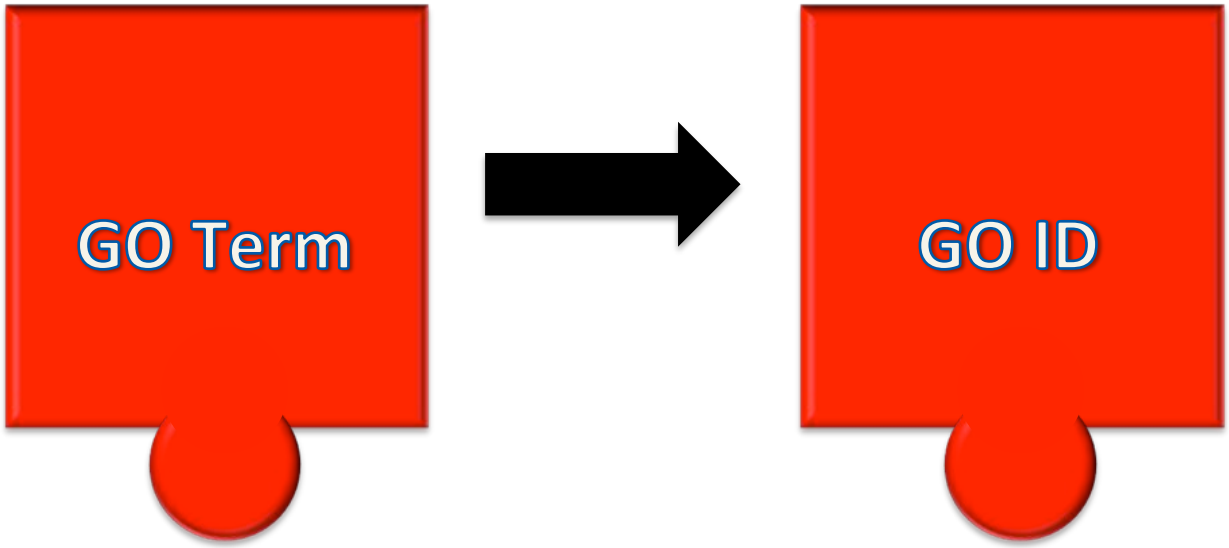


Components of an Annotation:

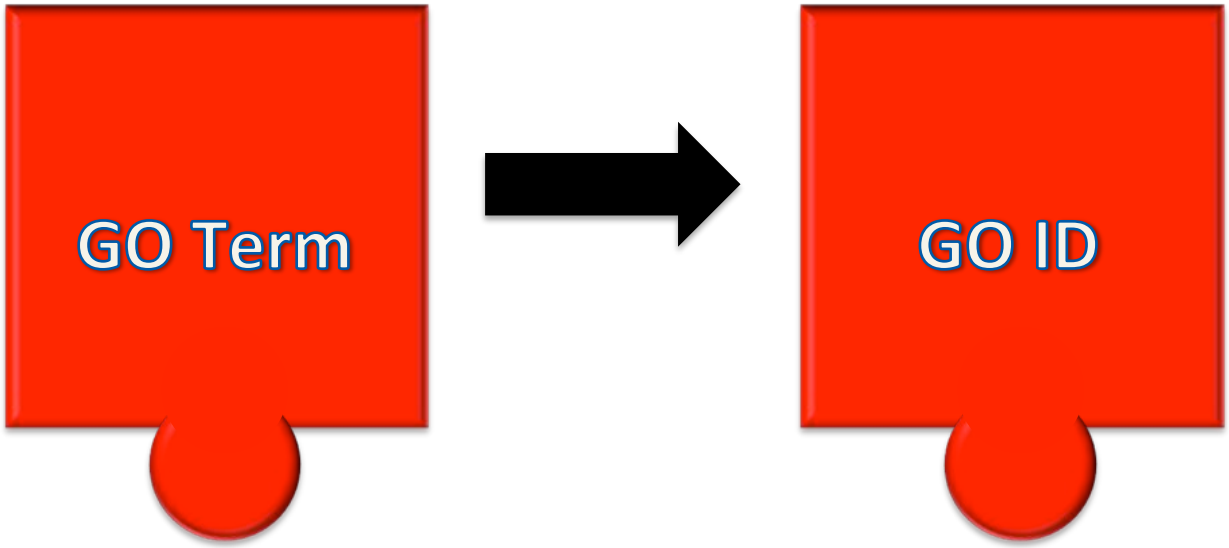
GO Term

Components of an Annotation:

GO Term

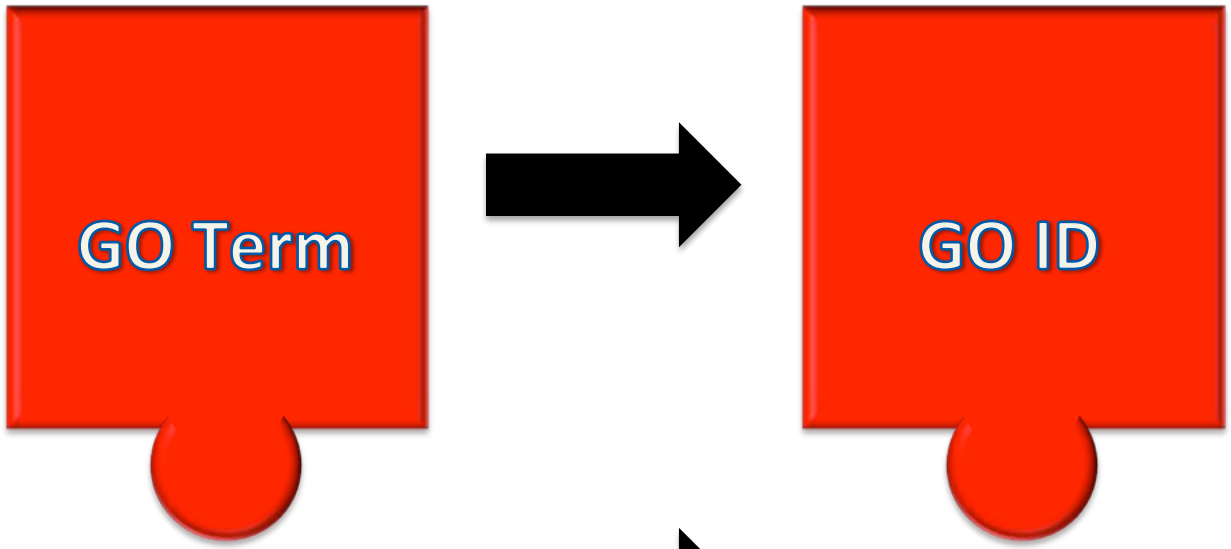


Components of an Annotation:
GO Term



- Find the GO ID using a Gene Ontology browser
  - QuickGO: http://www.ebi.ac.uk/QuickGO/
  - AmiGO: http://amigo.geneontology.org/amigo
  - GONUTS: http://gowiki.tamu.edu

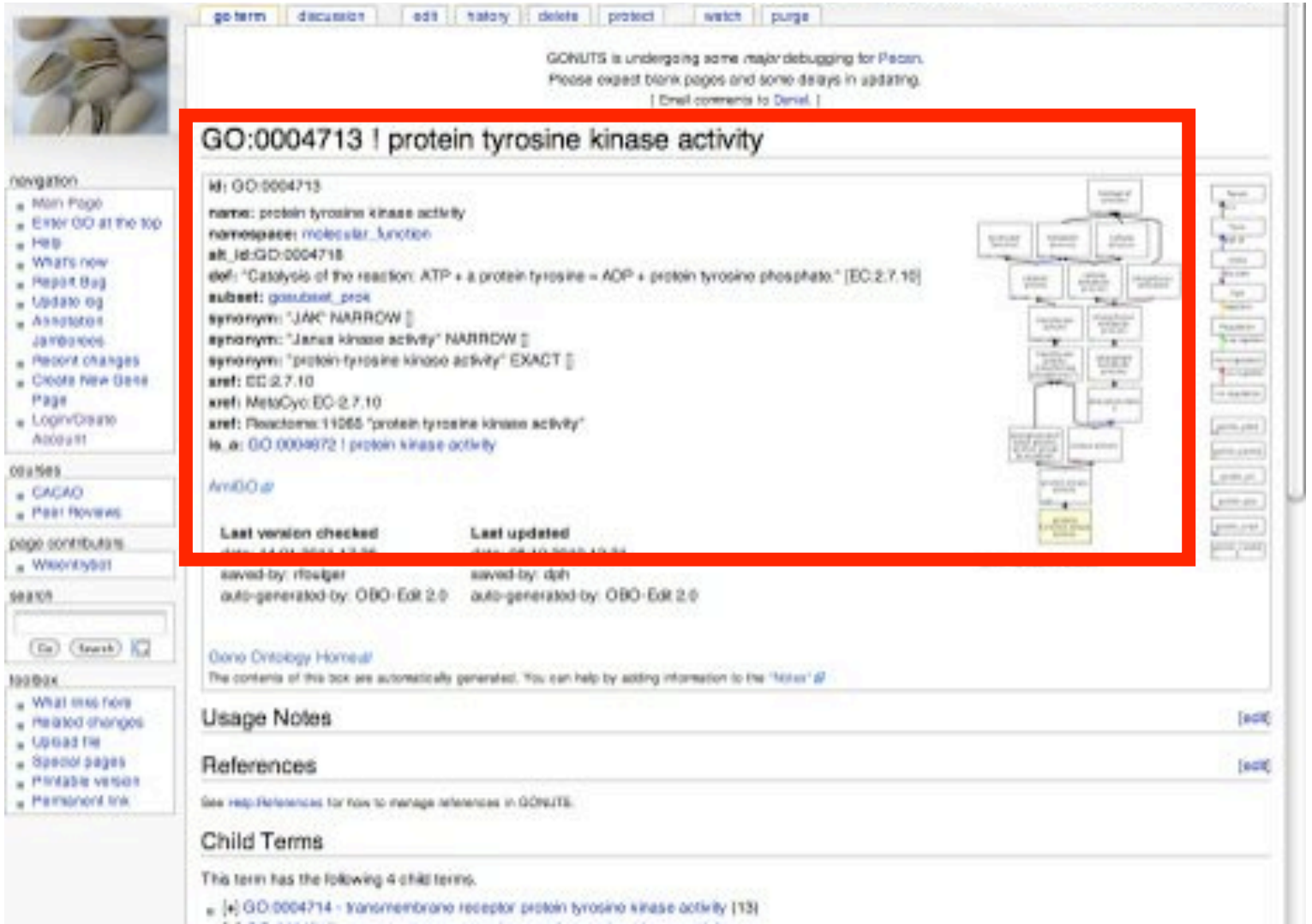
Components of an Annotation:
GO Term



protein monoubiquitination → GO:0006513

host cell endoplasmic reticulum → GO:0044165

Components of an Annotation:
GO Term



Components of an Annotation:
GO Term

GO:0004713 ! protein tyrosine kinase activity

id: GO:0004713
name: protein tyrosine kinase activity
namespace: molecular_function
alt_id:GO:0004718
def: "Catalysis of the reaction: ATP + a protein tyrosine = ADP + protein tyrosine phosphate." [EC:2.7.10]
subset: gosubset_prok
synonym: "JAK" NARROW []
synonym: "Janus kinase activity" NARROW []
synonym: "protein-tyrosine kinase activity" EXACT []
xref: EC:2.7.10
xref: MetaCyc:EC-2.7.10
xref: Reactome:11065 "protein tyrosine kinase activity"
is_a: GO:0004672 ! protein kinase activity

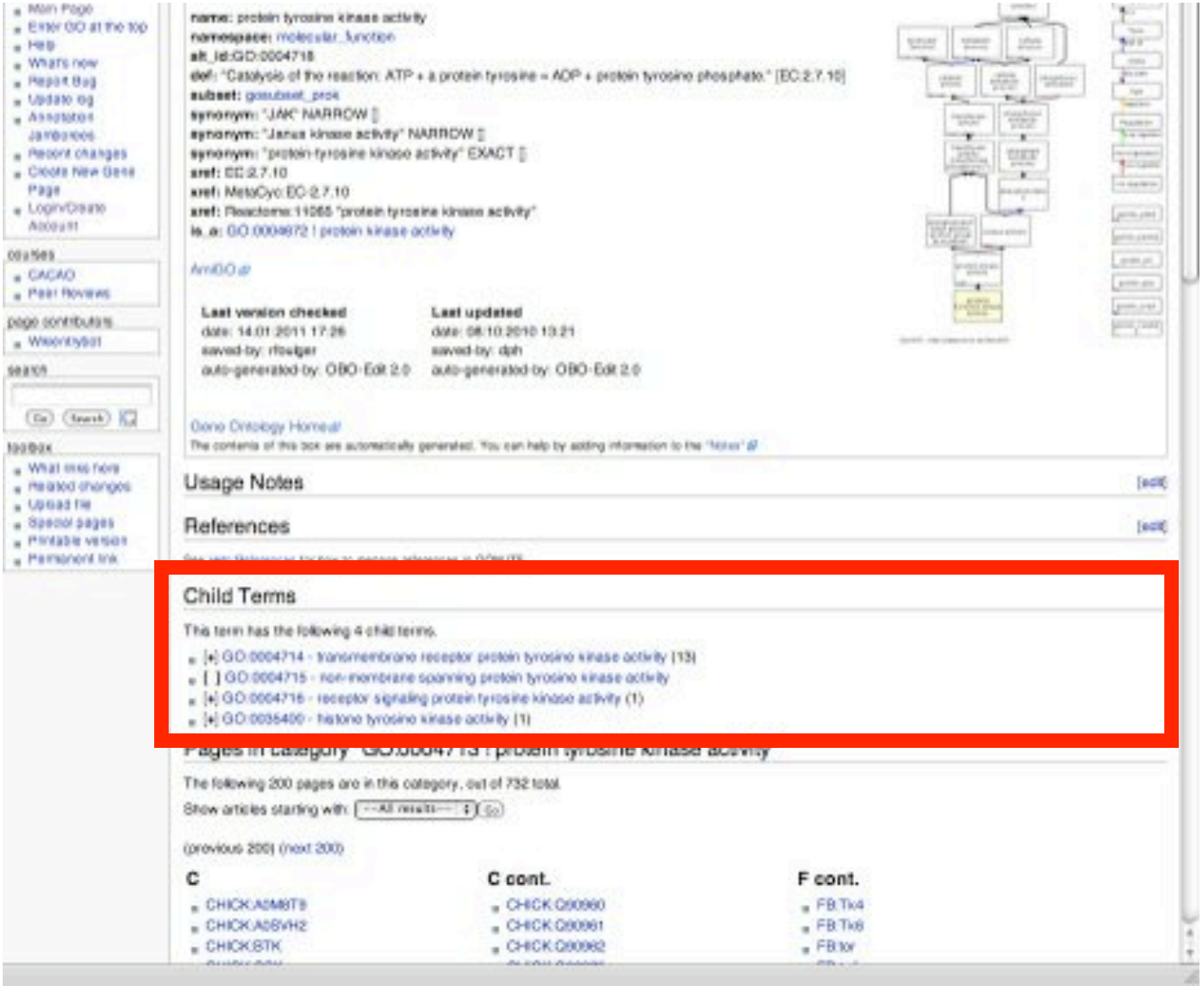
AmiGO

Last version checked	Last updated
date: 14.01.2011 17:26	date: 08.10.2010 13:21
saved by: rhodger	saved by: dph
auto-generated by: OBO-Edit 2.0	auto-generated by: OBO-Edit 2.0

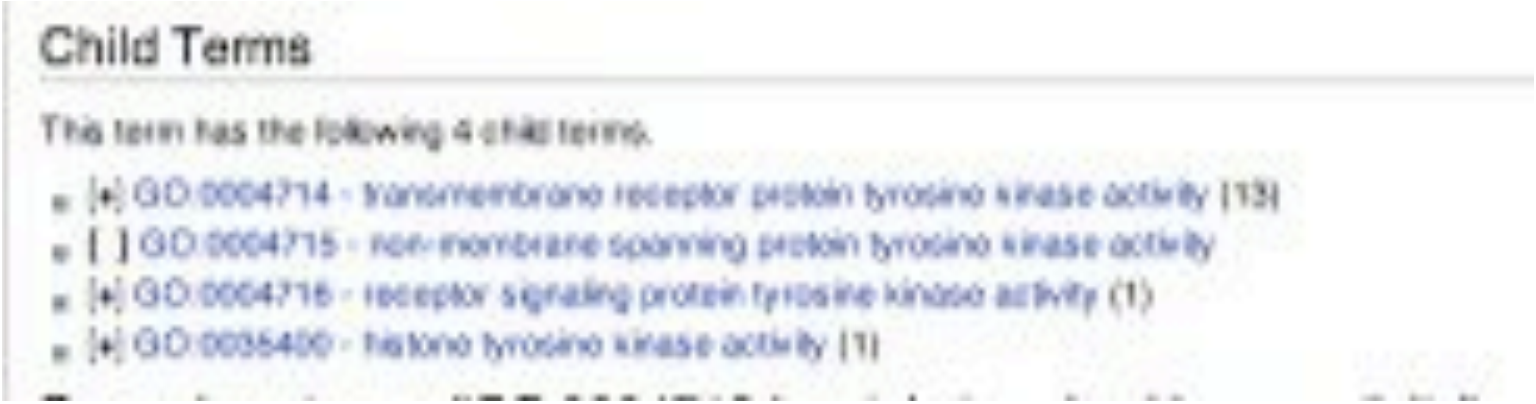
Gene Ontology Home

The contents of this box are automatically generated. You can help by adding information to the "Notes"

Components of an Annotation:
GO Term



Components of an Annotation:
GO Term

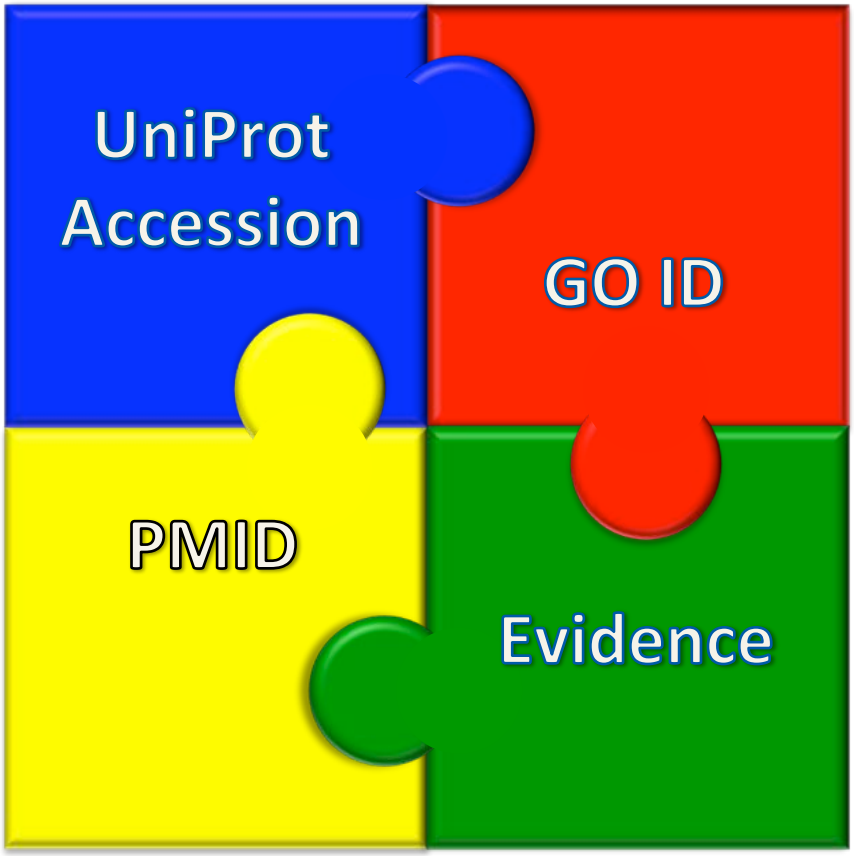


- Pick the most specific term the paper supports
  - Too vague = wrong
  - Too specific = wrong
- Ask for help. You may need to create a new term
  - "New Term Requests" count as annotation credit for your grade

Components of an Annotation:
GO Term
Strategies

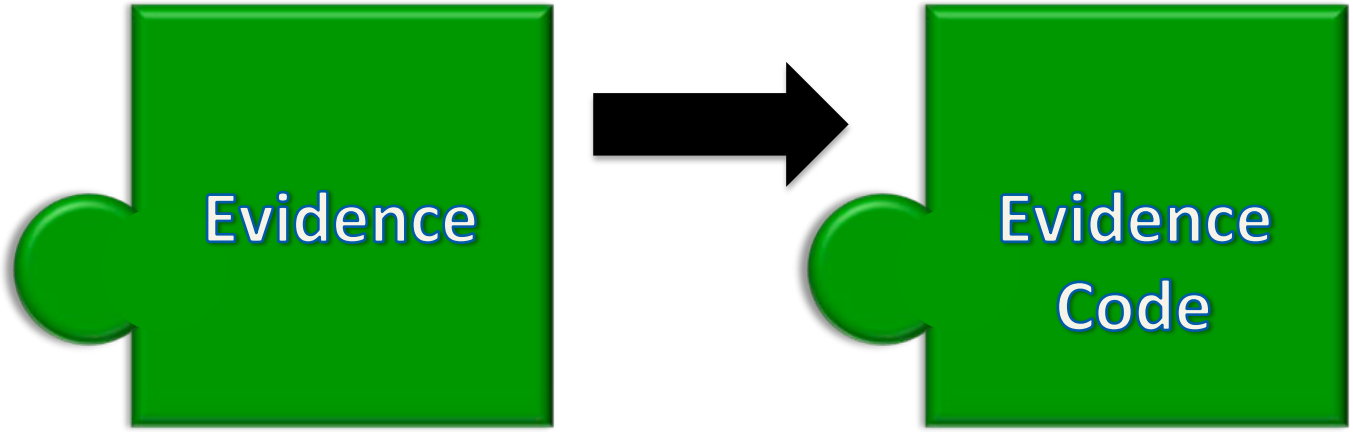
- Search for a keyword and browse the ontology for the right term
  - Look at the parents, children, and relatives
  - Use Google, Wikipedia etc. to find synonyms or alternative search terms
  - In GONUTS only search "Category" namespace if you get too many hits
- Look at terms suggested by others for your protein
  - Computational with the IEA evidence code
  - Curators with TAS or IC
- Look at terms used for homologous proteins in model organisms

Components of an Annotation:
What We've Covered

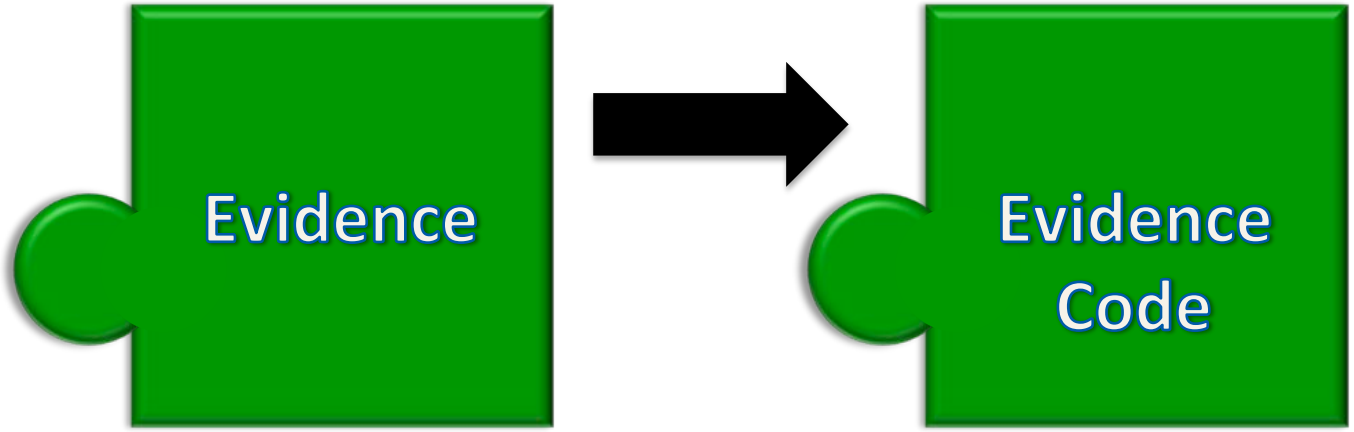


Components of an Annotation:
Evidence

Components of an Annotation:
Evidence

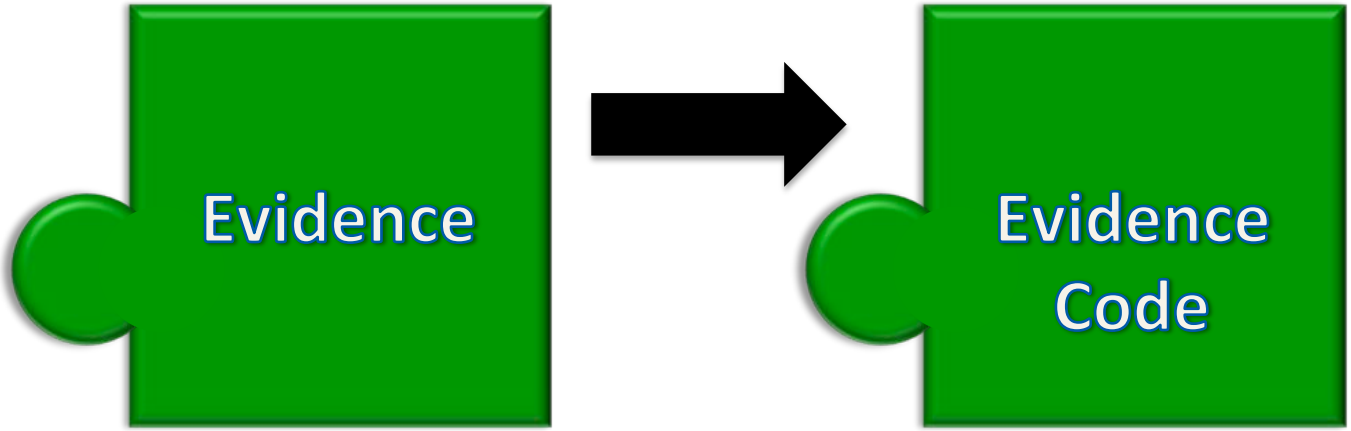


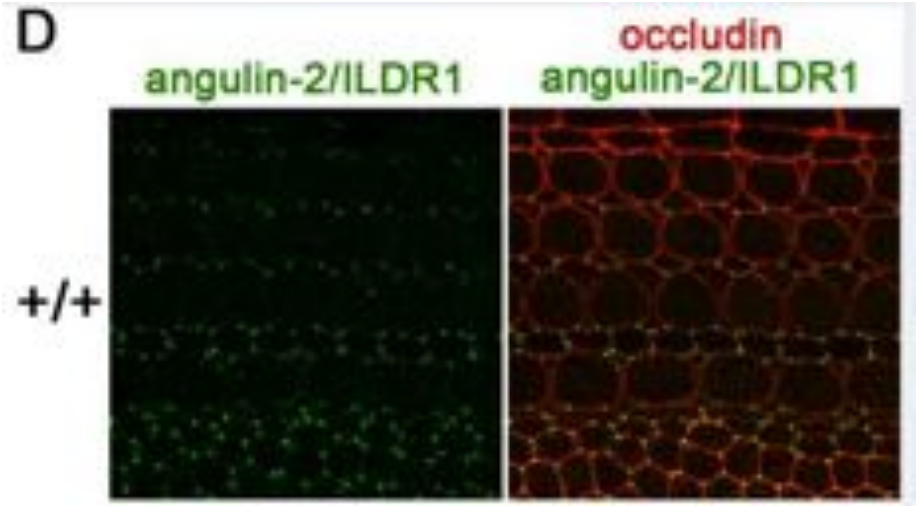
Components of an Annotation:
Evidence



- Evidence codes describe the type of work or analysis done by the authors

Components of an Annotation:
Evidence

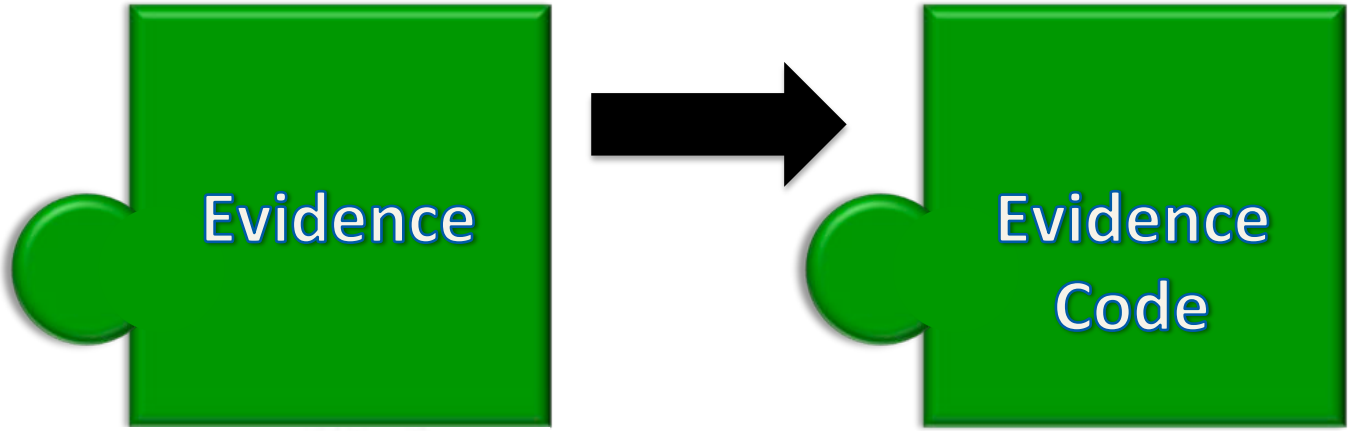


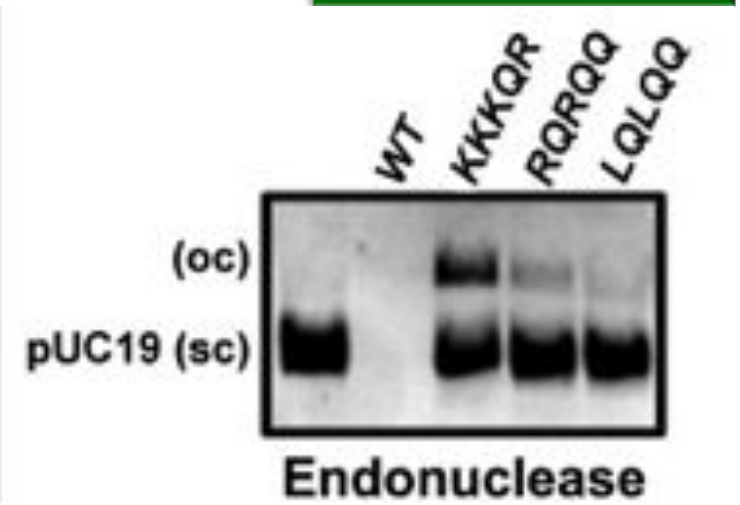


IDA

PMID:20167799

Components of an Annotation:
Evidence





IMP

PMID:25822906

Components of an Annotation:

Evidence

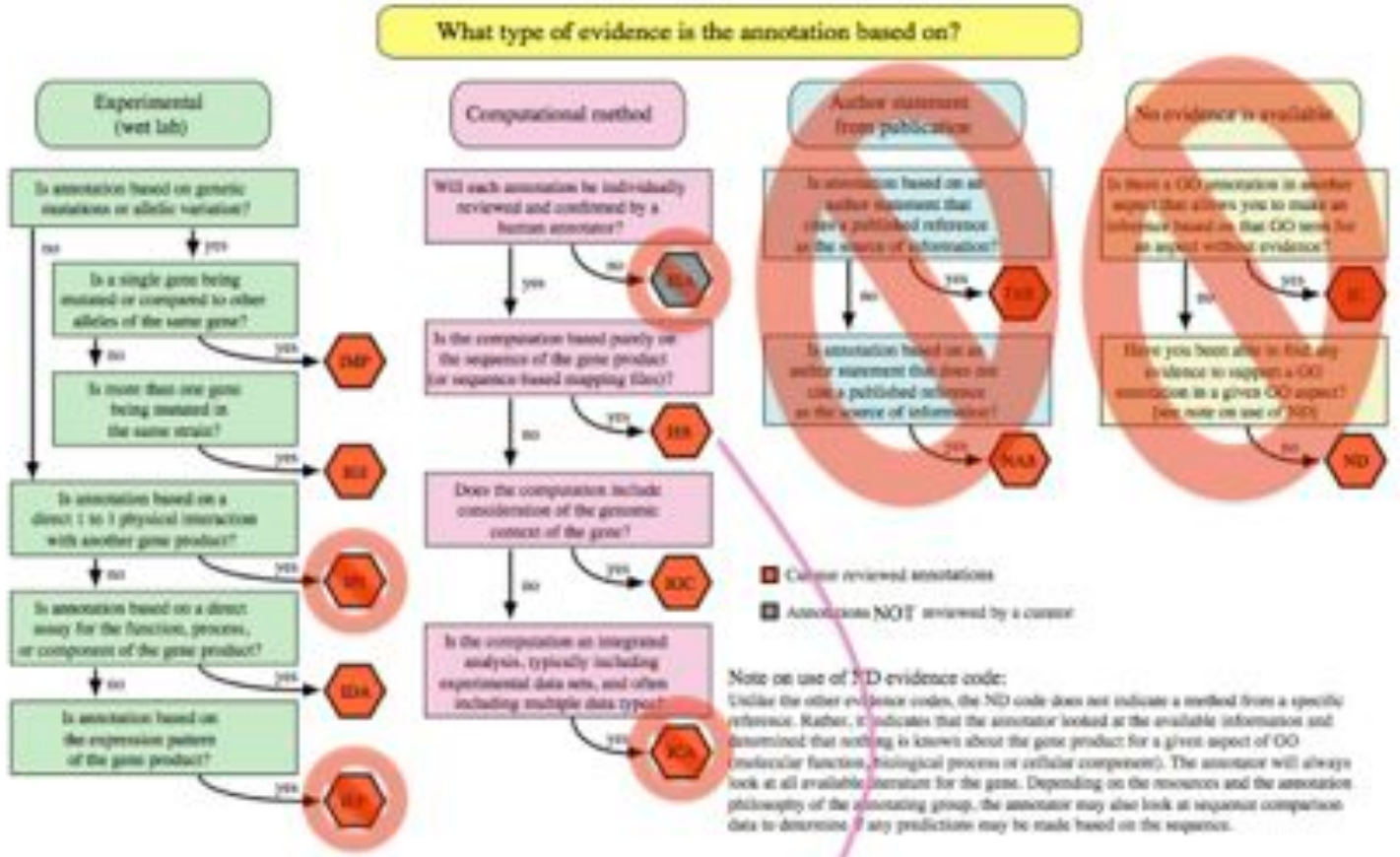
- **IDA: Inferred from Direct Assay**
- **IMP: Inferred from Mutant Phenotype**
  - NOT just for mutations! Includes inferred from inhibition in vivo by drugs, RNAi, etc.
- IGI: Inferred from Genetic Interaction
- ISO: Inferred from Sequence Orthology
- ISA: Inferred from Sequence Alignment
- ISM: Inferred from Sequence Model
- IGC: Inferred from Genomic Context

- Expert biocurators get to use others, but we restrict them for CACAO.

- http://gowiki.tamu.edu/wiki/index.php/evidence_codes

Components of an Annotation:

Evidence



ALLOWED CODES FOR ALL CACAO STUDENTS:

1. **IDA: Inferred from Direct Assay**
2. **IMP: Inferred from Mutant Phenotype**
3. **IGI: Inferred from Genetic Interaction** - requires with/from field to be filled in
4. **ISO: Inferred from Sequence Orthology** - requires with/from field to be filled in
5. **ISA: Inferred from Sequence Alignment** - requires with/from field to be filled in
6. **ISM: Inferred from Sequence Model** - requires with/from field to be filled in
7. **IGC: Inferred from Genomic Context**

Use one of these three codes (ISO, ISA, ISM) if the Decision Tree points you to ISS

Components of an Annotation:

Evidence

Some Evidence Types Require More Information

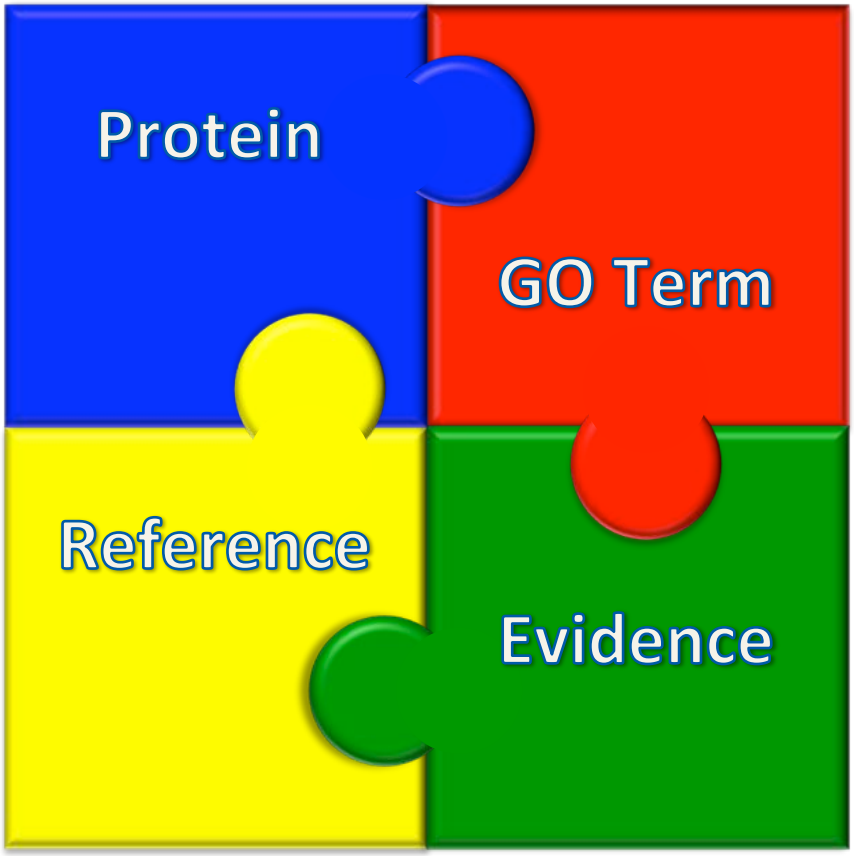
- *With/from*
- Evidence from sequence comparison
  - *With* the protein accession for the protein you are comparing to
    - "Match Protein" = comparison protein
    - "Match Protein" must have experimental annotation to the same GO term
- Evidence from computational analysis
  - *With* the reference for the analysis tool
- Evidence from genetic interaction
  - *With* the other gene(s) your protein is interacting with

- Ask for help if you suspect this applies to your annotation

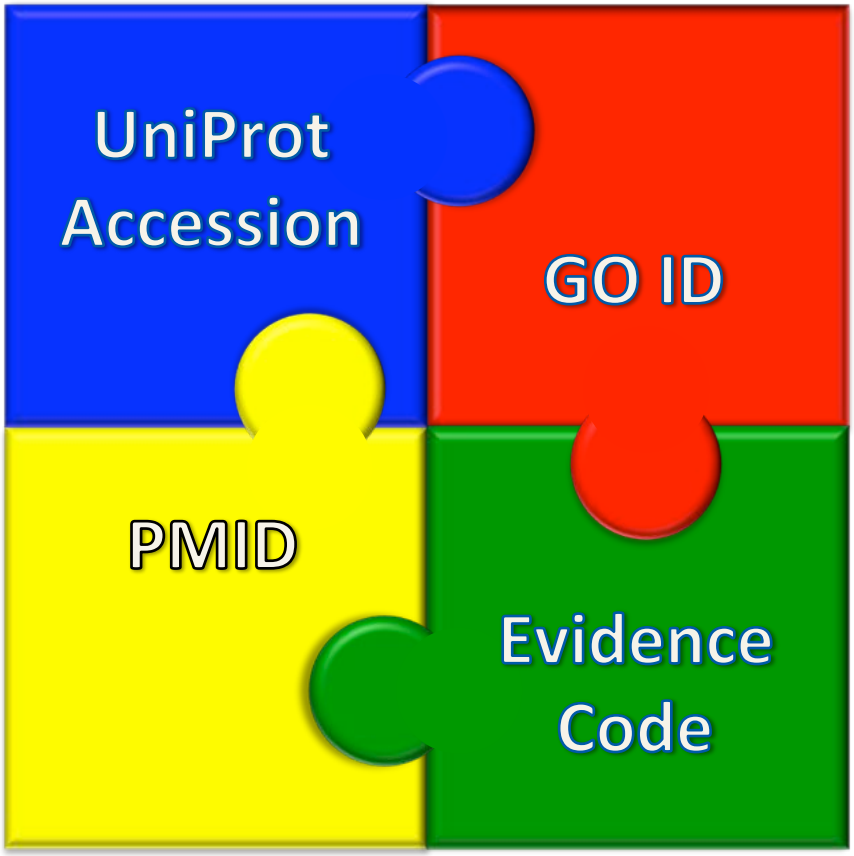
Components of an Annotation:
Evidence

- Picking the right evidence code is important
- Print & Use the evidence code decision tree
  - http://gowiki.tamu.edu/wiki/images/3/32/CACAO_decisiontree.pdf
- Use the evidence code guidelines at the GO consortium website:
  - http://www.geneontology.org/GO.evidence.shtml
- Discuss!

Components of an Annotation:
What You Need

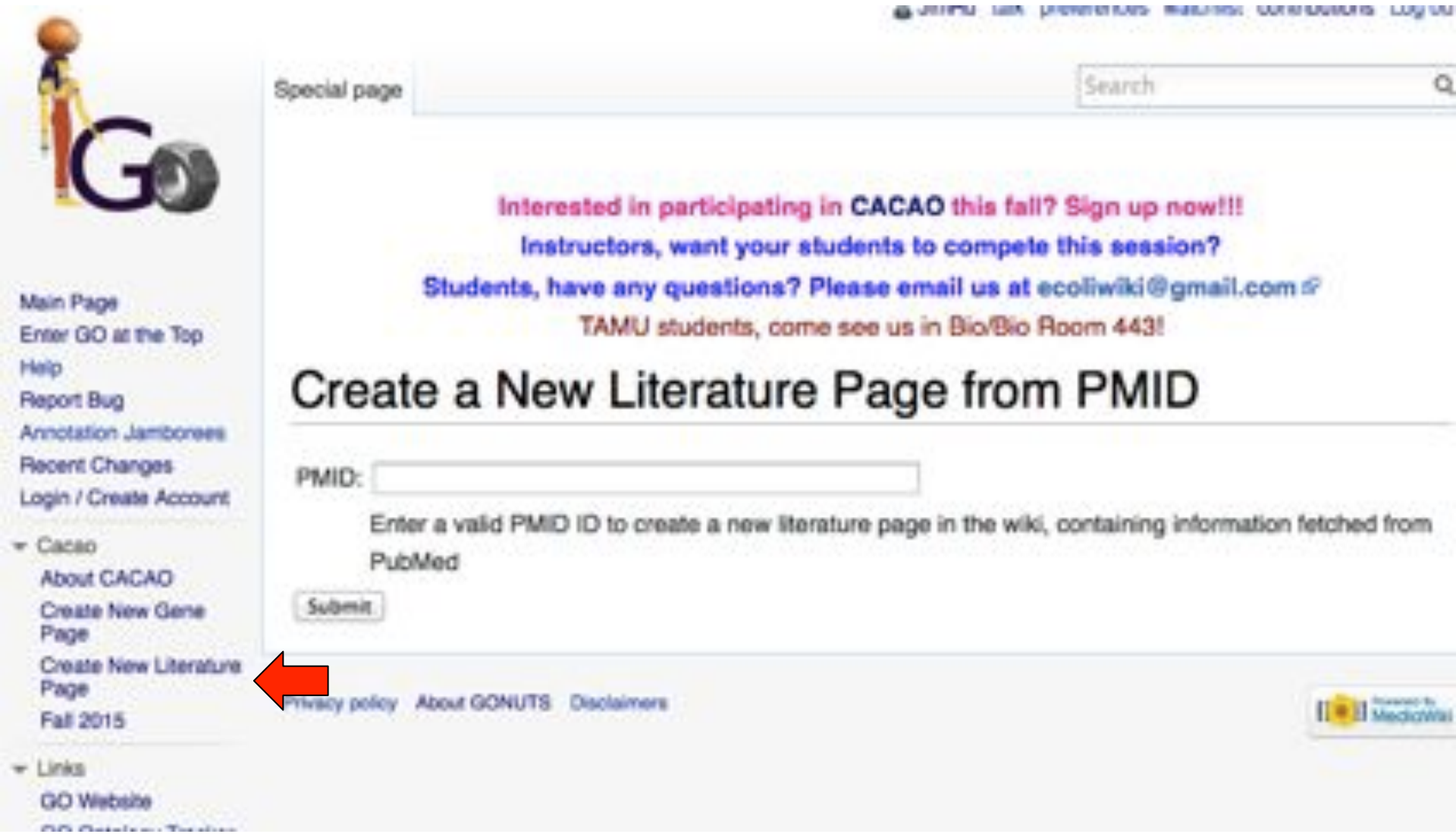


Components of an Annotation:
What You Find



GO ANNOTATION ON GONUTS

Create/Find a page for your paper

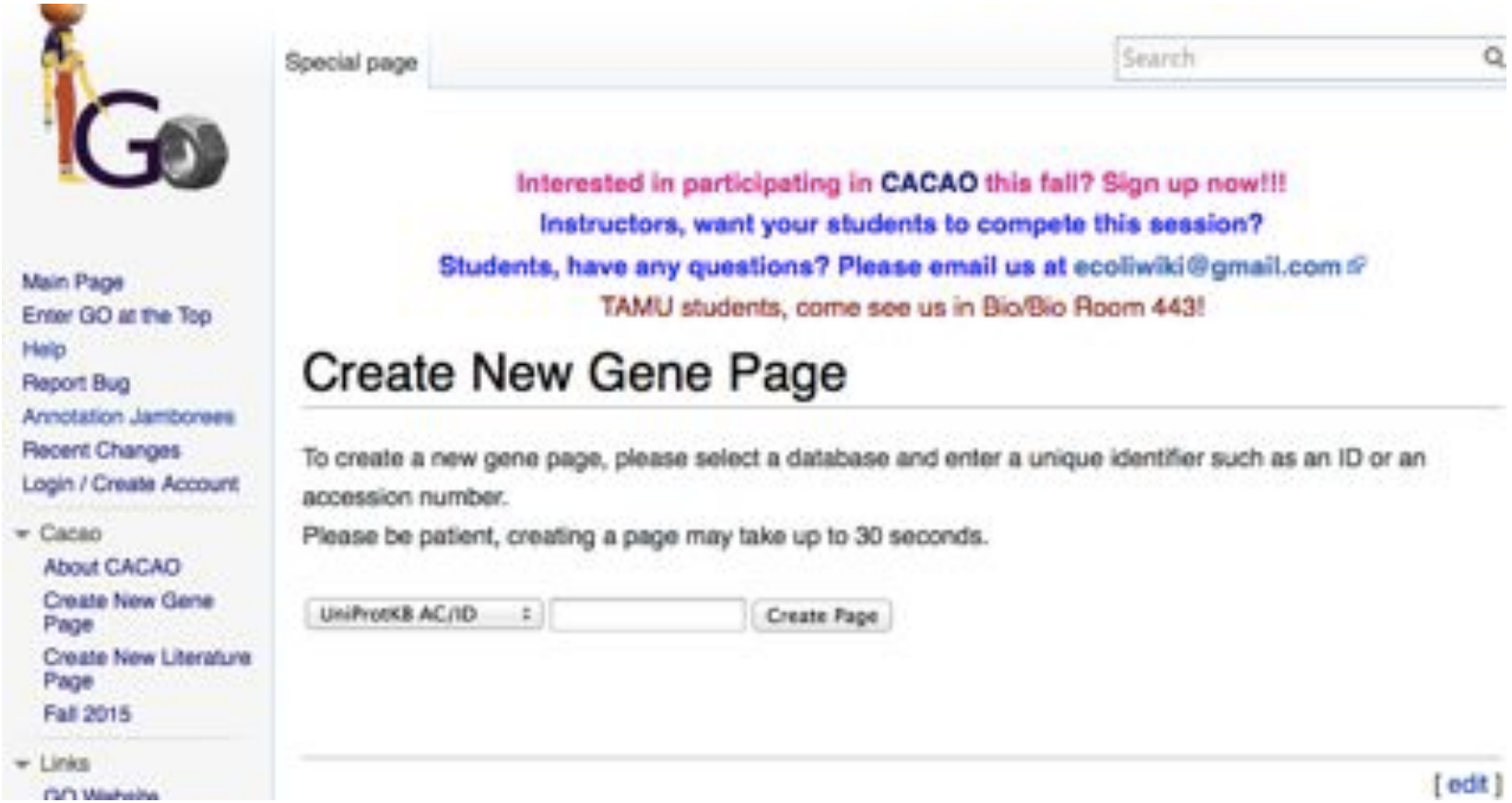


PMID Pages

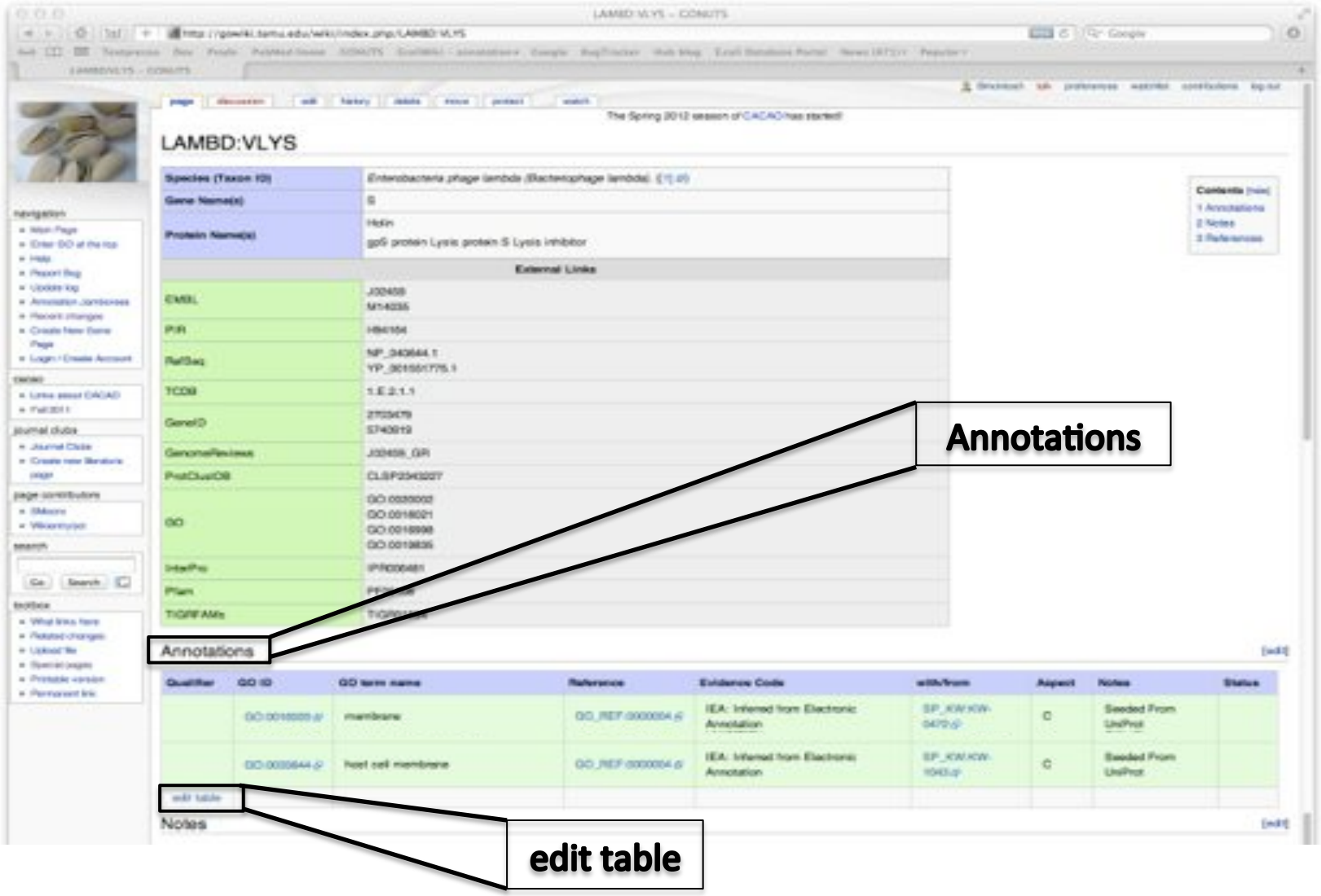
Abstract and full text links are automatically created

Anywhere the reference is used in an annotation is shown

Create/Find a page for your protein



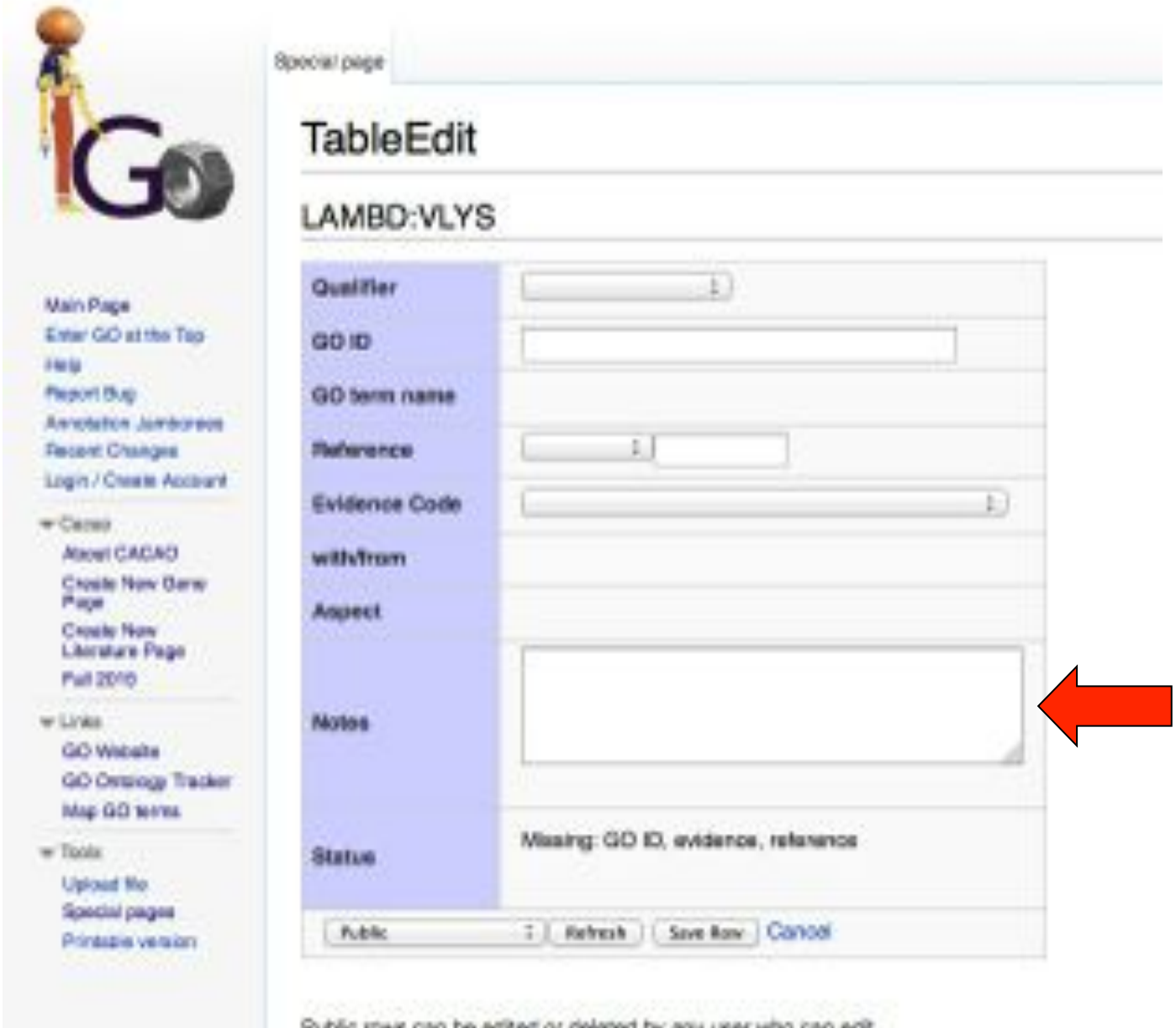
Protein Pages in GONUTS



Entering/Editing Annotations

Evidence Pull-Down Menu

Note Required for CACAO



Note Required for CACAO

- Notes **must** include:
  - Figure(s) and/or table(s) you used to choose the Evidence Code
    - You can narrow this down to panels if needed
  - Organism (strain if applicable)
  - Protein name/Gene name as paper refers to it
  - Protein name/Gene name as UniProt refers to it (if different)

Note Required for CACAO

- Notes **must** include:
  - Figure(s) and/or table(s) you used to choose the Evidence Code
    - You can narrow this down to panels if needed
  - Organism (strain if applicable)
  - Protein name/Gene name as paper refers to it
  - Protein name/Gene name as UniProt refers to it (if different)
- **Annotation will be marked "Unacceptable" if the above are not present**

Note Required for CACAO

- Notes **must** include:
  - Figure(s) and/or table(s) you used to choose the Evidence Code
    - You can narrow this down to panels if needed
  - Organism (strain if applicable)
  - Protein name/Gene name as paper refers to it
  - Protein name/Gene name as UniProt refers to it (if different)
- Notes should include
  - Explanation of methods if not obvious
- Notes might include
  - Anything else you need to tell us to support your annotation

Misc. CACAO advice

- Getting help is not cheating!
  - Talk to your teammates
  - Ask us questions
  - Talk to other professors
  - Email authors of papers
- Please complain if things don't work
  - Much of the custom software for CACAO was written by molecular biologists, not software engineers!
  - ecoliwiki@gmail.com

Example Papers

- Powerpoints and pdfs for different examples can be found at: http://gowiki.tamu.edu/wiki/index.php/Category:CACAO_training
  - Component annotation based on immunofluorescence in *B. subtilis*
  - Function annotation based on a study of *E. coli* Topoisomerase IV
  - More coming…

Possible homework

- Find annotations made by professional biocurators
  - Read the paper
  - Discuss what you would put in the note
- Examine past CACAO annotations
  - Discuss whether the annotations make sense
- Pick some papers that you might annotate
  - In class discussion of how you made the choice and why you think it works for GO annotation