Communications

02 July 2019 Batch-1 feedback call

Just for the interest of Arathi, here is one example that has discrepancies: In PMC5920331 Group1 has much fewer annotation on MIS\_GP and [MIS\_GP][ALL].

In addition, we’d like to clarify and discuss following things we noticed:

* PMC5708618, Additional green light generated with LEDs did not influence the growth of Cucumis sativus L. (Cucurbitaceae; cucumber seedlings).
  + Group 1 annotate “Cucurbitaceae” as WT\_OG, which we think it should be a correct organism name although “Cucurbitaceae” is a family name in NCBI taxonomy.
* PMC3892176, Group 1 has fewer [MIS\_GP][ALL] which includes kinase.
  + Because of [ALL], it may have a big impact during post process to match annotations in an article.
* PMC5315555, Group 2 annotate the same gene-disease relationship examples as both “YGD” and “NGD”.
  + If annotators are not sure about the relationship, tag “AMB” can be used for ambiguous examples

With some points from Arathi:

1. Where do we a draw a line on when a term is to be considered as a broad term? Examples- proteins, kinases, cytokines, transcription regulators etc.

2. Should we annotate term like “CsbHLH18”, which is a composite mention of both the gene/protein as well as the source of the organism [Citrus sinensis in this case]?

Summary after call:

Q. Where do we a draw a line on when a term is to be considered as a broad term?

A. Very broad terms like proteins, genes, enzymes, receptors are out of scope.

Terms like kinases, cytokines, transcription regulators, transcription factors, amylolytic enzymes, antioxidant enzymes, E-box binding proteins, glucan hydrolytic enzymes, protease, Neurotrophic factors, bZIPs, transcription factors, RTKs, tyrosine kinases, receptor tyrosine kinase, phosphoproteins, RTK, tyrosine kinase, photoreceptors are to be annotated.

Q. Should we annotate term like “CsbHLH18”, which is a composite mention of both the gene/protein as well as the source of the organism [Citrus sinensis in this case]?

A. YES

Q. Can we annotate taxonomical families like “Asteraceae”, “Lamiaceae” as organism?

A. YES

Q. Can Just the strain name without the organism detail be considered as entity type “organism”. For example-There are instances when the strain of an organism is mentioned along with the organism name like “P. aeruginosa PAO1” but in subsequent instances, only the strain name is mentioned in the article like “PAO1”.

A. Stand alone strain name should not be annotated. Organism information is mandatory.

Q. Should we annotate term like “CsbHLH18”, which is a composite mention of both the gene/protein as well as the source of the organism [Citrus sinensis in this case]

A. YES

Q. Is there a preferred way of capturing term like “P. aeruginosa lung infections”- Is it preferable to keep them together as a disease or split into organism and disease entities?

A. “P. aeruginosa lung infections” should be taken as one entity.

Q. There could be cases where a broad term like “infection” or “acute illness” is auto annotated- Do we retain such terms?

A. In case of broad terms that are already annotated in EuropePMC, if the annotation is correct it should be marked as such. However, if there are any missed instances of such terms subsequently in the article, the missed annotation comment need not be tagged.

13 August 2019 Batch feedback

Thanks for your email. We had a meeting for reviewing some of the annotations. As usual, the annotations are really good. We notice that for the PMC5731848, the gene-disease relationship annotations are slightly different between annotators.

Here are two examples:

"We introduced single or combined *Rad54* and *Parp-1* inactivating germline mutations in *Ptc1* heterozygous mice, a well-characterized model of medulloblastoma, the most common malignant pediatric brain tumor.”

In this example, the authors introduced "Rad54 and Parp-1 inactivating germline mutations” in “Ptc1 mice”. There is no relationship mentioned in this case.

“Our study reveals that combined inactivation of *Rad54* and *Parp-1* causes a marked growth delay culminating in perinatallethality, providing for the first time evidence of synthetic lethal interactions between *Rad54* and *Parp-1 in vivo*.”

Precisely, “growth delay” is a phenotype rather than a disease. So if we treat it as phenotype, there is no relationship in this sentence.

In PMC5070310, in one case, “HIV” has been annotated as disease but as it is a virus it should be organism.

No further feedback for batch-3 and batch-4.