

Tuesday afternoon, 02/28/2012

Mike Cherry:

Project is not just to create an ontology but also creating annotations. We want to merge the two aspects: Ontol Devs work with annotators, etc.

Want to remove barriers so that ontology, annotation, infrastructure are more seamless.

Paul Thomas:

Organization of GO Central

Wrt to the aims of the grant:

Annotation productivity -> annotation streams

What are the roadblocks?

Questions about consistency, term usage

Annotation Integration and Ontology Review

Centralized QC system

Review of annotation corpus as a whole

Framework groups

Infrastructure – database, ontology logic

CAT (common annotation tool)

Users and Community - >users

Making sure we're serving our user community

Focus on specific ad hoc teams for specific deliverables within this overall structure

Mike Cherry:

Realize there are lots of different aspects to the resources we provide, but for this discussion we are talking about the higher level goals.

Questions

Eva Huala:

Will certain people be assigned certain areas and will that be their full-time responsibility?

Mike Cherry:

Yes and no; We need lots of help; we need task-oriented teams. Will try to have more shorter term, defined tasks.

Not necessarily a long-term commitment, but may be focused on a particular area of biology, for example.

The teams should be more cross-expertise.

There will be cases where people will be assigned (in conjunction with GO Council).

Documentation, for example, is hard and will likely need to have the work assigned.

GO top will be mandating some participation at some level.

GO PIs: Mike, Judy, Suzi, Paul, Paul

GO Council: Monte, Eva, Rolf, Jim, Rex

Paul Thomas:

Again, ad hoc groups.

There will be people dedicated to particular areas; but will bring together people across all areas to solve problems, as needed.

Ruth Lovering:

List tasks and groups that are doing it (recognition)

Great to have teams with short-term projects, but it'd be nice to feel that there is a recognition of the groups, that we're not calling on the same groups and people all the time. We need to recognize a good job and be mindful about who's doing what.

Mike Cherry and Paul Sternberg:

In some cases, publications would be appropriate recognition.

Ruth Lovering:

Balance annotating with working group participation.

Adds ontology development to newsletter, for example.

Paul Sternberg: Over the next two weeks, will refine more specific topics with the manager's groups. These topics will evolve as time goes on.

Suzi Lewis: next time we meet we have things to actually check off

Next Meeting

Sept. UK? Maybe Edinburgh (Chris sponsor)

Suzi Lewis: Working groups may want to get together (5 people); if necessary then GOC will try to find the funds.
USE the GO conference line.

Jim Hu: make a checklist/manual on host to a GO meeting
Spring 2013 in Texas?

Tanya Berardini: for now proceed as usual until told otherwise.

Mike Cherry: yes

Soybase:

PAINT is good but there are lots of empty nodes in the tree. Getting those nodes filled computationally is important; how to do that is non-trivial. Would a useful activity be a guided assessment of use cases? Can we have a test case to see when and how we can automate clade curation? Don't see that an army of curators could fill in the leaf nodes. That's a difficult task. Could evaluate fleshing out of the tree.

Paul Thomas:

In pilot, took half a day on average per family. There are 7000 families. It is doable this way, but can we start to look at automated rules now? What could we get? Could curators at least evaluate the automated annotations?

Pascale Gaudet:

Triangulation method when there are two annotations does not work well.

Suzi Lewis:

Only experimental evidence annotation is pulled from the db. If we can get a closer loop between the annotation tool and PAINT annotations, then that would help.

Try to pick where there are a lot of ISSs for a family vs pairwise homology comparisons?

David Hill: growth factor annotated to Golgi; we don't want that probably.

Kimberly Van Auken to Paul Thomas:

What is best bang for the buck? How do you select the families

Paul Sternberg:

some just have IEAs...

Paul Thomas:

growth factors across animals... Process is always hard

Transcription factor families as a subset of processes.

Cellular processes vs organismal processes.

David Hill:

Having poorly annotated families may be a way of prioritizing.

Judy Blake:

Can solicit ideas about what tasks/curation to prioritize in the next 3-6 months.

Pascale Gaudet:

Report that how many non-redundant each db gained

Make sure the literature annotation first.

Judy Blake:

expressivity vs??

Pascale Gaudet:

shows amigo labs phylotree: amigo-sven.princeton.edu/cig-bin/amigo/phylotree

List of PANTHER families, number of genes per species, number of genes with experimental annotations.

Tool indicates when a family is annotated and when many new annotations are added, amongst other things.

All seem last updated in 2010 dec.

Prioritize families according to how many annotations they have.

Comprehensive annotation data is key for prioritizing PAINT.

Paul Thomas:

Found that there are a lot of annotations in the database; you can pick something randomly and find a family that can be annotated.

David Hill:

For prioritizing, can we create a mathematical approach to choosing which families to annotate?

Judy Blake:

run Panther tighter to get small families?

Paul Thomas: You can limit yourself to curation of the subtrees.

Tanya Berardini:

Then you may not be able to use experimental annotations across the whole tree.

Paul Sternberg:

For compound eye development, that's essentially organ development.

Cindy Krieger:

Can you load the IEAs in PAINT just to review, not necessarily to propagate?

Kimberly Van Auken:

Change curation paradigm from organism to topic/gene set?

Judy Blake:

What does this gene do (not in mouse or fly)

Already being done to some extent by human, mouse, and rat.

Evolution of trust between annotation groups would more readily allow for this type of annotation paradigm.

Peter D'Eustachio:

get groups together to do topic

Have a new paradigm of PAINT annotation – talk to experts. Would provide outreach and expert knowledge. Use this approach to get

breadth and depth. Piloting this approach would allow for estimates of what we'd get from this.

Paul Thomas:

See comparative table in paper - InterPro2GO vs Compara vs PAINTE

Ruth Lovering:

Should we be focusing on Compara set (since we know they have something)

Judy Blake:

For mouse, only have experimental data for a fraction of the genome. There are sets of genes that we know something about.

Julie Park:

Mouse genome, 8000-9000 genes with experimental annotations, another 13000 that don't. To be GO complete should be a goal, but that's in opposition to expressivity? How does that fit together?

Judy Blake:

Experimental annotations, ISS annotations, IEA annotations. Expressivity refers to cell type, time of action, etc.

Suzi Lews:

Is it the case that if we're doing expressive annotations, then we're not covering the genome in enough breadth? We are trying to address both.

Judy Blake:

We need to have ways of prioritizing annotation across organisms.

Paul Sternberg:

Another comment on expressivity. There may be contexts in which the expressivity is really important to understand the biology of that gene product.