Sesssion: Vision for GO annotation process (Feb 26, 2012)
Minutes: Rama, Karen

Implementation of standards

Emily (see pdf)

Slide 1
essential GO annotation activities for all groups
req for GAF 2.0 format
format & update req's for gp2protein, gp2rna files
transition to ECO ids
if group cannot comply with regular release schedule, responsibility reverts to GOC
This document will be eventually published on the GOC website
Jim: question about IDs: must provide a gp2protein if you use a non-UniProt ID, must include all proteins, even those not annotated

Suzi: this is a minimum standard to make sure GOC has enough info to be able to maintain later. If some of the tasks can't be done, then GOC will take the responsibility to convert it.

** Judy: all single heritable traits not mapped to sequence coords, but with GO annotations, how do we keep these in the annotation set? GAF file can include these, but these would not be in gp2protein file. Request to update documentation to be explicit

** Susan Tweedie: suggestion to have a third file to be explicit about which IDs do not have sequence.
Ben: having a file with nulls allows cross checking that an ID in GAF is known to not have seq, versus an error in not including it in gp2protein.

ACTION ITEM: Decide on whether there should be an additional file for annotated gps that do not have sequence. Whatever decision is made, update documentation to be clear on all files: GAF, gp2protein, gp2rna, possible third gp file for.

Val & Paul S: use of these genetic GO annots in enrichment => further discussion about whether or not these are useful in enrichment, but cut off for now as not the focus of this meeting.

Rama: use of ECO ids, what will we do if annots are provide using old GOev IDs. Suzi responds that this is mostly documentation for new groups coming in. Either traditional GOev codes or ECO fine.

Emily: GAF 2.0 does not require ECO; but going forwards we will be using ECO.
Suzi: wants it to be a controlled field.
Rex: will we be discussing how to implement this?
Suzi: yes. ____ UniProt is switching to using ECO ids within this year.
David: Is there funding for ECO? is it going to be sustained.
Chris: We can always take over ECO.

Slide 2:
Responsibilities for active GOC groups
responsive to requests for annotation corrections (from curators, external users, QC checks)
curators should regularly attend annot calls and GOC meetings

Responsibilities for authoritative annotation source for a species
- integrate manual and IEA annots from external sources on a monthly basis. This should include data from annot_extension field (col 16) when provided.

Suzi: purpose here is to be explicit about what is expected if a group is authoritative source, or if a group does not wish to meet these standards, then GOC can be responsible group.

** Tanya: question about req to take IEAs from UniPRot, TAIR already runs their own InterPro pipeline. Susan: added comment about whether intent is to provide all vs non-redundant set. Judy: intent is to provide non-redundant set, we can work out full details about what comprises non-redundant set, but we are just agreeing to load external annotations at this point. Mike: many of these InterPro annots are good, intent is to make sure that all groups are including these. Tanya: suggestion to centralize filtering of non-redundant at GOC, so that MODs just provide all, and there is one consistent standard for what is included in IEA set. David: common annotation tool: each group receives the official set from GOC, rather than each group submitting to GOC. Mike: software dev may be able to do comp btw manual and IEA. Suzi: promise that prototype for CAT will be available in 1 year. Val: comment about IEAs derived from exp ev from her organism, these are not providing additional info for e.g. pombe

ACTION ITEM: Work out details of exactly what is meant by non-redundant annotations and what exactly is required for groups to incorporate, or absorbing external annotations

slide 3

Expectations of GO NIH funded curators
Annotation activities (supported by central GO CAT)
Annot files 1x/month
Transition to more expressive annot formats

Create highly informative annots (many aspects)
Contribute to discustions on annot expressivity, e.g. LEGO
Make annots for all species described in a paper
Do PANTHER annots via PAINT
Other activities
Participate in dev of CAT
Mentor new GO annot groups
Participate in review of annots contributed by external community experts

** Non-NIH groups: Ruth: Ruth: Varsha and I curate other species when we come across those...shd we continue? Absolutely (GO ruth!)
Rolf: we have one FTE paid by GOC. highly unlikely we will use the CAF. I won't commit to something that I am not infrmed about.
Suzi: we are talking about giving feedback on CAF. You don't have to commit to development of this tool.
Judy: we want everyone to understand what our goals are and our efforts are.
Mike: we'll talk later, we'll need to make sure understandings are clear about how NIH groups can coordinate.

Judy: many MOD groups have 1 or a few GOC-funded curators, more curators that are not GOC-funded; this workflow applies only to GOC-funded curators.

Pascale: What is relative weight of importance between "annotation activities" and "other activities". Top activities seem relevant to all GO curators, while "other" activities seem specific to the GOC-funded curators? Judy: we'll need to develop priorities.

ACTION ITEM: Work out details on which of these activities are top priority.

Annotation submission (Suzi, Paul T., Chris and Emily)

Paul T: Enhanced annotation expressivity presentation

Discrepancy btw richness of ontology and simplicity of annotations
ontology is rich
annotations are very simple – each aspect has single implied relation to GO class
MF, has_function
BP, involved_in
CC, located_in
Enable curators to use GO to express rich biological statements from literature
Extensions to enable annotation expressivitiy
MF: molecular specificity, e.g. kinase acts on specific target
Relations (and GO classes) for creating statements using >1 GO term

BP regulation and molecular mechanism, e.g. func of gp1 reg's func of gp2
Between "compound MFs": e.g. kinase has 2 coupled activities: ligand binding & kinase activity
Statements across ontologies: e.g. gp1 executes MF kinase activity (while located in) CC cytoplasm
Example
Abstract
UV-induced RNAPII Ub'n & degradation are important DNA damage responses
Nedd4 acts as an E3 that functions in this
Disjoint annotations
Use of "RNAPII binding" does not fully represent the richness of the knowledge that RNAPII is being targeted for degradation
What we'd like to express (picture)
Nedd4 is a specificity factor that Ub's RNAPII and shuts down txn in response to UV
Suspiciously like a pathway diagram
Shows that specific MF is nested within a BP

Ruth: comment that you cannot make this complex annot from that single paper Paul T started with

Rolf: question about how much curator time is involved in doing this? Peter: Quantitative comments about how much of the genome needs to be covered in this way for it to be useful. At 1/25th, not so useful, at 20% already useful. Shouldn't we structure it quickly before lots of work is lost. Karen: frequently the parts of the MF and the BP involved in txn are not in the same paper. David: Why are we doing this: for ourselves or for benefit to our users? Rex: Identifying which parts are demonstrated vs inferred. Ben: Only feasible if it doesn't cost much to curator time. Pascale: Potential to help make sensible annots, not need to find multiple disjoint annots. Judy: chain of evidence Peter: we have been spending 12 years making these diagrams, not trivial, probably uninferrable from the disjoint GO annots. Requires expertise from curators. ___ : software solution needed. Difficult to compose these from disjoint GO annots, difficult for user to find info. Karen: Often will be common stories where we could generate "template stories" then curator could indicate which specific piece is supported by specific evidence statement, at very little extra cost in curator time. ____: need to connect which MF involved in which BP, this is what end user is really looking for. Rama: How does this affect enrichment?

Rex Nelson(plant): from the plant perspective, we often come across things that work differently in a given species, how does this work. David: If the ontology is actually too restrictive, or not right for all species, then we need to know. Paul S: this is the right way to go, we need to investigate if this can work and how to make it possible. Rolf A: limited amt of money with more scrutiny going forwards, must convince funding agency that this is best bang for bet, cannot do it all, must decide where to invest the money.

Judy: this meeting is to put all the ideas out there and then we can decide the overall priorities.

Suzi: have not covered reorganization of management, will have more working groups with limited duration and targeted goal. Kimberly will become a manager & Huaiyu Mi also.

Action Item: Do a pilot to see how long it takes, what are the bottle necks etc.