Monday 2nd October

Welcome, overview, vision, introductions

Paul Thomas

GO handbook presentation

Presentations are on the Google drive:
https://drive.google.com/drive/folders/0B7bEr6HANSlGSHY3c2JqNEs2ZUk

- The Gene Ontology and the meaning of biological function (Paul T)
- Translating research data into Gene Ontology annotations (Pascale)
- Gene Ontology - annotation extensions (Ruth)

Ontology Issues and Updates

Update on MF refactoring

Pascale & Paul T
Background: MF ontology has drifted away from terms describing overall function, toward specific/granular aspects like binding. Was difficult for terms created later (nuclear receptors with protein binding component, TFs, etc) to be wholly described. Aim of restructuring was to make MF more navigable at the higher levels. Rebasing MF classifications to reflect true biology rather than constructs formed by eg enzyme commission.

Clarification: No classes will be deleted

Hijacked molecular function added for virus receptor-interactions - also for taxon checks, separate it out from “normal”
Molecular Carrier Activity? Not yet fixed, electron transfer activity - move out under ox-red: call for comments.

New top level classes - the grouping mechanism of protein function.
Molecular role/system component function. To give a more biological view of MF. Example of oxido-reductase (molecular) vs. electron transfer (role).
Put some of the important functions at the “top of the ontology” to make them easier to find.
Try to reflect names used in literature rather than very complicated standardized names. E.g. use “nuclear receptor” rather than “RNA polymerase II……..ligand-activated….blah”

Annotation Issues and Updates

Qualifiers/Relation issues

Biological Process Relations: Use of Qualifiers in Legacy Annotations

Pascale

Proposal: We will apply the general qualifier to all legacy annotations. Each group can provide more specific qualifiers if they have a mechanism to distinguish. This is a general blanket application of a generic qualifier.

- These qualifiers are necessary for the GO-CAM models

Causally_upstream_of_or_involved_in

Paul T: discussion at Corvallis was that some groups, because of the way they annotate, are comfortable making the more strong statement part_of

Sylvain, Helen: training will be key

- Chris M: can/should build into the GO-CAM training, then focus on GAF, GPAD output later

Val: Don't we need to have start/ends of processes to be able to determine if something is upstream?

Regulation_of will be problematic, depends on annotation strategies employed by different groups

- Some groups, when unsure, annotated to the term, other groups when unsure annotated to the 'regulation of' term
- Some clean-up will be required, can look at products annotated to both a term + regulation_of the term

Sabrina: Will need to impose a reasonable deadline (ex. 2 months)

    - First discussed at Barcelona

Pascale: can look at PAINT... EXP + IBA gives stronger evidence for part_of and/or involved_in

ACTION ITEM: Pascale

- We will implement the new qualifier "causally upstream of or involved in", with the more specific children "causally upstream of" and "involved in"
- Groups will have a certain amount of time to include the new qualifier in their GAF file
- If by the deadline, some groups have not provided the qualifier, the general one will be added by GOC database as a post-process on the GAF files
- Documentation/training:
    - Definition of causally_upstream required
    - There will be training for curators to explain how to use these qualifiers (ideally this will be through the Noctua training)
    - Will need to discuss at annotation calls (curators bring questions to calls)

  Strategies to identify annotations that meant 'involved in':

- Make use of PAINT

- Need to make tools available to everyone - can Mary provide the regulates AND process annotation results for all groups?

Regulates relations (Kimberly)

****Point of this presentation is to agree on how we want to do this.****

Adding new qualifiers for the relation between a gene/gene product and a GO term.

    - Corvallis discussion:* What should the default relation be? How will we handle regulation? Use a relation, involved in regulation of, or use the precomposed regulation term?

Ruth: There are now 3 ways to say the same thing: - involved_in_regulation_of X - involved_in X regulation - involved_in BP regulates(X) For annotation purposes and for our users we want one. DOS: Good point. These are semantically identical, but I agree we need to find a way to only have one: by convention for classic GO annotation and by filtering the output of inference for noctua output. Proposal: If a named regulation class exists: involved_in X regulation ...if not: involved_in {some BP} regulates(X) NOTE: If there was an annotation in Noctua such as 'regulation of' 'very specific term' and there was no term as 'regulation of very specific term' the GOC pipeline would create the annotation to the parent term: 'regulation of less specific term'. Q: Is this implemented ?

3 ways to say the same thing (semantically equivalent):

1. Gene product involved_in X regulation
2. Gene product involved_in Biological_process (root) col16 regulates (another process)
3. Gene product involved_in_regulation_of X

#1 is the most familiar, and will provide best enrichment with *existing* tools (which don't use qualifiers)
http://noctua.berkeleybop.org/editor/graph/gomodel:59cde02c00000053
lunch break, then resume discussion…

DECISION:
This is the way we are going to do for GAF/GPAD: #1 Gene product involved_in X regulation
Will use default 'involved in'

- Regulates relation in Noctua will create regulation_of terms on the fly
- **ACTION ITEM**: Kimberly: Document decision that for assertions of type "Gene product involved_in X regulation", we will use default 'involved in'

Multiple qualifiers for an annotation (Huaiyu)

Impetus for this topic: Huaiyu saw some annotations in AmiGO (one annotation with contributes_to, one with NOT), asked some questions

Huaiyu: Need proper rules for using multiple qualifiers to avoid ambiguity
Chris M: need to distinguish between 'qualifiers' like 'NOT' and 'relations'

ACTION ITEM: Kimberly need some guidelines for these multiple qualifiers

ACTION ITEM: Kimberly need to come up with a proposal for the next GO mtg for '**contributes_to**'

CC component annotation guidelines (Kimberly)

CC component annotation: what does it mean ? Kimberly to do 1 proposal (out of 3 alternatives) :

1. where the protein is active
2. two different meanings: enables or the right RO (part:of, ie just found there)
3. part_of (low information value!) ++ colocalizes with

'Enables_activity_in' v. 'part_of'...'found_in'

Scenarios in the literature encountered by curators:

- CC where known to be active
- Novel CC, but no reason to think it's not active there
- CC where known to not be active
- CC where not known if active or not

Possible options:

1. Only annotate where product is active using more explicit 'enables_activity_in'
2. Only annotate where product is active and continue using 'part_of'
3. Annotate both where active using 'enables_activity_in' and inactive using 'part_of'
4. Continue active and inactive using 'part_of'

Winner: 3. Annotate both where active using 'enables_activity_in' and inactive using 'part_of'

Colocalizes_with

"Transiently or peripherally associated" - leaves much room for interpretation, meaning/usage not consistent

For cases where binding to/ association with a complex has been shown and complex content is not clear, ask Complex Portal to define content of complex, then curate to that entity, as described. You can request a complex term and annotate as 'binding with':
www.ebi.ac.uk/complexportal → request complex

ACTION ITEM: Kimberly

- Survey groups to see how many want to express BOTH active location and just 'incidental' locations
- Define exactly the relation name: is the relation 'part of' be used for CC? Suggestions: 'localises in'; 'found in'
- Legacy annotations: will default to 'part of', but groups can move to 'enables_activity_in'. We'll also use the same strategies as for BP to assign 'enables_activity_in'
- DEADLINE: Proposal: By the next GOC meeting every group will have added the qualifier (have to edit your GAF production pipeline) or a default value will be assigned by GOC (can change them in the future)
- Tell Tony what the qualifiers are to put in P2GO
- Each group needs to update their GAF-generation pipeline to integrate the qualifier. For each group that does not provide any information in the qualifier, these will be automatically filled in by the GOC to the more general qualifier ("part of" for CC/ "causally upstream of or within" for BP)

ACTION ITEM: Kimberly

- Review EXP annotations using co-localizes with
- Consider creating new GO terms if necessary (for eg, 'peripheral to ribosome')
- Consider creating more precise qualifiers, for eg 'transiently localizes with' or 'dynamically localizes' …

Protein complexes

GP is part_of complex is part_of cellular location
→ both part_of's are effectively 'enables_activity_in' as we only use CC for active location
This was not discussed: we need a group to formulate the issue and propose a solution

ACTION ITEM: Birgit, Kimberly, Harold
Separate working group for protein complex representation in Noctua: Formulate the issue; make a proposal (see above)

Report from signaling workshop - Kimberly

See slides

Ontology Group update

David - See slides

Noctua update

Getting Noctua ready for production (Kimberly & Seth)

Attribution - curators that curate for different groups need to contact Seth to get that set up (if not done already)

- Can see all this is the users.yaml file
- Check to make sure your entry is correct

Annotation extensions - how many and which? Noctua reasoner often puts many more than curators would put in, but they are all 'correct'

ACTION ITEM: Chris: List of remaining tickets that need to be closed for Noctua to be 'production-ready'
ACTION ITEM: Chris:: document where to slurp up GPADs (request from Tony, but certainly useful for other people as well)

Review wish list for Noctua from Corvallis:

- Provide ways for users to recover and digest GO-CAM units (Gene Ontology-based Causal Activity Model). Ideas include rule-based generations of text statements from model, cytoscape view of network described, etc.
- I (Seth?) believe that this is a more exploratory and open project. We now offer an RDF endpoint, in addition to the APIs we already have, as well as the proposed feeding from AmiGO. If this is still open in the future, it may be very good Hackathon material. Needs more discussion.
- **ACTION ITEM:** Create a working group to discuss the development of end-user tools for displaying Noctua models David, Chris are interested PIs Need to prioritize the development of end-user tools for displaying Noctua models

- **ACTION ITEM:** Pascale + QC working group think about how we'll do QC for Noctua models

There are tickets for all following issues. We now need to prioritize.

- ECO codes available for use in Noctua should show how they map up to a classic GO code, and there should be an alert for curators when they are using a code that does NOT map up to a classic code
- There is no ticket for such functionality. This would likely have to be written into Minerva or as a pre/post check on (some) server.
- PRO IDs for use in Noctua
- https://github.com/geneontology/noctua/issues/429
- https://github.com/geneontology/noctua/issues/122
- Fix GPAD export from Noctua
- https://github.com/geneontology/noctua/issues/418
- Add a SPARTA workbench
- https://github.com/geneontology/noctua/issues/465
- Working group discussion of evidence on complex Noctua models (see Day 3 for AIs)

Noctua table form demo

Chris

Project updates

Report on transcription work/Noctua templates: Astrid GREEKC consortium

Slides: https://drive.google.com/drive/folders/0B7bEr6HANSlGSHY3c2JqNEs2ZUk

Main job is to agree how we are going to represent data
GREEKC is still defining what they need to represent; ongoing discussion

Report of Reactome-GO connection: David H/Peter D

Tuesday 3rd October

SynGO meeting report: Paul T, Ruth

Trying to annotate genes known to be in the synapse
SynGO ontology development = VU team + GOC team (DavidOS)
SynGO Annotation practices

- Different model for annotation than any of us have used before; some elements of community curation (encouraged them to learn from Val)
- There are some training/communication issues; need better guidelines and free-er exchange of information between sides
- Annotation quality has improved over time, so gave them go-ahead to start releasing

Lesson learned: need to be involved earlier with external groups trying to start GO
Idea: have yearly training opportunity for new groups?

ACTION ITEM: Write up recommendations and guidelines for next time we engage an external group of experts

Alliance of Genome Resources - report to GOC: PIs + Stacia

Alliance = 6 MODs + GO
1.0 public release - 20 Oct 2017

- Gene pages (overview, seq. Feature viewer, GO ribbon, orthology set, disease associations)
- Disease pages (overview, disease associations)
- Only experimental evidence is included in GO ribbon
- Stage.alliancegenome.org -- two weeks of testing; report bugs using “Submit bugs here” link

Are there documents for using the Alliance site?

- Not yet. Too new.
- More MODs may be added to Alliance

Data update frequency?

- Known need; frequency hasn’t been determined yet; versioning system also to be implemented

Alliance funding?

- Institute that funds Alliance may change, but there should be funding
    - Key message here = “Dont worry”

Data Commons Pilot Consortium (NIH)

- Sharing resources in the Cloud
- Subset of MODs have funding for small projects (6 months)

Documentation in GO (Kimberly/Seth/Pascale)

More and more documentation being moved to GitHub
For ontology development, used GitHub to write up the documentation (.md files)
http://go-ontology.readthedocs.io/en/latest
All the documentation will be linked through the website as it evolves. Documentation is for GO editors

- If you go to wiki for documentation, please e-mail Kimberly so we can prioritize what documents to move

If your comment or concern or request isn't in one of the GO github trackers, then it doesn't exist and won't be addressed. Get your comments/needs into github.

GitHub overview

- Add your ticket to helpdesk if you don't know where it should go
- Go-ontology tracker: new term requests, missing relations, fix definitions, taxon constraints, ontology pipeline issue
- Go-annotation tracker: requests about annotations to review, InterPro2GO feedback, PAINT feedback, requests for docs
- Noctua-models tracker: integrated with noctua; Will be merged with go-annotation tracker
- Other repos of interest: go-site, amigo, paint, helpdesk
- Kimberly takes care of annotation tracker; David takes care of ontology tracker
    - Adds assignees to tasks, adds labels, proposes prioritization
- Some unused labels may get removed. Let Pascale know if you are still using an underused label
- Every user should have a GitHub account
- Every contributing group should have a contact person
- If assigned on annotation tracker unassign yourselves once dealt with your part. Last assignee to close ticket

Decision on what axis of classification labels will be used for and then clean up the labels and only have labels for that one means of classification.

Centralization of InterPro2GO annotations (Paul T)

Proposal (follow-up from Geneva 2016):

- GO database pulls directly from InterPro2GO for UniProt Reference Proteomes
- MOD identifier is used as primary gene identifier (human mapped to HGNC IDs)
- Annotations are given "contributed by" InterPro
- MODs pull from GO database, no need to maintain separate InterPro pipelines

Are there any objections?

Chris: 2 independent proposals
Kimberly: this is dependent on organisms having a reference proteome set;
Paul: UniProt identifiers can be used to define a set
Val: We filter everything that we already have experimental annotations for; we don't submit all InterPro2GO annotations into GO. What you see in PomBase will be different - same for Flybase
David: MGI has 'black list'.

ACTION ITEM: collect uber 'black list' and post it somewhere that everyone can access it
ACTION ITEM: Come up with some rule for filtering ahead of time - new GO rule for filtering InterPro2GO for GO pipeline - https://github.com/geneontology/go-site/issues/436
ACTION ITEM: put together working group of interested individuals, have a couple meetings, come up with guidelines by May 2018 GOC mtg
ACTION ITEM: Each of the databases should provide their QC pipeline to make sure nothing falls through the cracks

HTP guidelines (Helen)

Helen - 15 minutes Report on progress from HTP working group

- We already have HTP papers annotated using GO; some good, some bad
- Need to indicate that these come from a set that will contain false positives

- Draft Guidelines https://docs.google.com/document/d/1ScIeclAzUXMe-tU6n0lVfsSwMHpOeNb7uK8On9-iKXc/edit?usp=sharing%7CHTP
- Provision of new evidence code
    - New evidence codes are sitting in ECO tracker
    - Will send notifications to all interested parties
- Implementation & adding to guidelines
    - Curators ask yourselves: am i adding value or adding noise?
    - Help with choosing a good dataset with low false positives
    - Guidelines re: experiment codes
    - Case studies
    - ECO codes should be available soon
- Exercise restraint! Don't overuse these HTP ECO codes; not about getting a lot of HTP datasets, but about getting high quality HTP datasets
- Don't use for ISS transfer as provenance will be lost (until QC checks/rules are established)
- Are we going to go back to check that these older HTP datasets meet these guidelines?
    - We will contact the groups that are on the list of the datasets that the group has reviewed

ACTION ITEM: Metrics group

At the next GOC we'd like to discuss metrics

ACTION ITEM: QC rule for ISS: HTP-types will not be supported for ISS

- For PAINT : To be discussed

Transcription annotations decision tree

Ruth Action Items Corvallis 2017/06

- David OS to create transcription regulator activity (proposed by Paul T)
- Proposed changes to decision tree in Corvallis:
    - Simplified from previous version. Essentially a choice between 'regulating transcription by RNA polymerase II' or 'regulating gene expression'
    - Annotation 5 = contributes_to sequence-specific DNA binding

David: when people do enrichment, they don't drop contributes_to (ie., pay attention to qualifiers), so all those proteins will come down as 'DNA binding' Action item: replace 'annotation 5' with ISS annotation (if DNA binding domain) or contributes to annotation 5 with ISS annotation if no DNA binding domain and domains to suggest coactivator

- Annotation 3 = nuclear chromatin

Not everyone comfortable with this (ex., Stacia, David) Shouldn't it be 'colocalizes_with'? It was previously agreed that the definition for nuclear chromatin: The ordered and organized complex of DNA, protein, and sometimes RNA, that forms the chromosome in the nucleus. Source:PMID:20404130 was to be applied. The proteins here include all associated proteins, not limited to just histones. With this statement the TFs are then contributing to chromatin, rather than binding to chromatin or colocalizing with it.

- Decision tree to be put up on GOC website.
- Should annotation 4 exist? (Ruth)

- Ruth will send out a survey for decision tree changes
-

ACTION ITEM: create transcription regulator activity:
https://github.com/geneontology/go-ontology/issues/13588

Astrid: chromatin contains/includes transcription factors (eukaryotes only)
Unresolved: prokaryotes: location of transcription factors? Certainly not at nuclear chromatin?
For eukaryotes: the decision is that transcription factors are *located at* nuclear chromatin (not part_of), and that they DO NOT bind chromatin (so don't include MF 'chromatin binding'
RUTH (or someone who understands the issue) PLEASE COMPLETE

ACTION ITEM: Ruth To be discussed in Lisbon, follow up at next meeting
ACTION ITEM: Annotation groups Review TF annotations to 'chromatin binding'; can these be transferred to 'nuclear chromatin'? How much of this can be done (semi)-automatically?

PAINT update (Huaiyu)

Huaiyu Action Items Corvallis 2017/06:

- Encourage discussion between PAINT curators and other annotators about terms not used for propagation
- Report how many annotations per species are used for annotation propagation; could even supply this number for propagation specifically to human genes
- Ruth and Huaiyu (others?) will discuss making use of groups that have already annotated specific gene lists to annotate the corresponding PAINT families.
- Smooth out the challenge mechanism to make it easier to do make and resolve

the challenges, identify terms that may be problematic and would benefit from consistency exercises and discussion.

- Get a list of families where terms have not been propagated (?) - Please check this one for clarity.

(added to github project board) Develop mechanism to trigger review of annotated PAINT families.

Citing GO (Paul T)

- Web page - temporary, not linked to anything currently -
    - Cite original paper + current update paper
- Tool providers: Action Item Corvallis 2017/06: someone to represent GOC and contact GO tool providers to display version information and state this information needs to be included in any subsequent publication. Plus list the GOC paper to reference, as well as the tool provider reference. See: https://github.com/geneontology/go-site/issues/359 and https://github.com/geneontology/go-site/issues/360

- Current page: http://www.geneontology.org/page/go-citation-policy
- Page that Paul has been working on: http://www.geneontology.org/page/citing-go

ACTION ITEM: Merge the two pages

Enrichment (Paul T & Suzi)

*Web page
*Paul Pavlidis…
*Data Commons work

GO_Slims

WHY creating slims - Intro to slim philosophy (Mary, Val, Suzi)

Suzi

Different types of slims

- For an overview of a genome
- For profiling a gene
- For a taxonomic group
- For curator sanity (and consistency)

Question: should GO maintain slim? For genomes ? For genes ? or should GO merely provide criteria and methods ?

David - do we want the overview and profile slims to be different?

- Val: Slimming tips Slimming tips

Includes Slim uses, Creating a biologically useful slim (complete coverage by aspect, biologically useful terms i.e sufficient granularity, avoiding single step process terms (i.e functions), different slims for different purposes) (Val: 20 mins)

- Whole genome slims -- good resource to plan coverage; what isn't mapped to slim.
    - Don't necessarily want big overlaps between terms
    - Run only 1 slim aspect at a time
    - To judge the slim it is useful to distinguish unannotated/unknown/unslimmed, to make sure that the results represent the primary annotation set

- Mary D For overviews, for particular taxa, for a particular area of biology

Use and maintenance of slims (Mary and Suzi)

Mary: algorithm - 15 minutes

HOW to make slims & Standard Metadata for slims (Chris)

- Metadata for GO slims
- Metadata isn't in the header, and it should be
- Encode a standard metadata for each subset
    - Proposed information: contact, URL, taxon, short name, full description, date first created, update protocol, archived/obsoleted
- There's a ticket in ontology tracker for keeping track of slims
- Goantislim_grouping slim
- SOP for new slims -- in the docs

Yaml format for creating slims/types of slims/target organisms

* https://github.com/geneontology/go-ontology/issues/12780
* https://github.com/geneontology/go-ontology/issues/14028
* https://github.com/geneontology/go-ontology/issues/12554

GO presentation of slims

Suzi: GO ribbon Group discussion: how should we present slims in GO: Multiple or one ? How many slims will be available? How will we represent multiple slims? (Pascale asks: what are our resources for this? what is the priority? which slims is GO responsible for?)

GO SLIMS TO DISCUSS/TO DECIDE

- (Suzi): We need to establish **criteria** for creating and maintaining slims - WHETHER we will create/maintain the slims
- We need to establish **guidelines** for creating and maintaining slims
- Which slims are useful? Who is using them?
- Using slims to prioritize areas to curate ?
- Chris MIAAGS: metadata about slims - TO DOCUMENT

Wednesday 4th October

Working Groups

List is at
https://docs.google.com/presentation/d/1AYS7IaeyrdzQcl42huyooh2y00kAj6gUtaVS5v0czjg/edit#slide=id.g25a0d9ffc9_0_0

Author Intent/Background Knowledge

Working group discussion

- Examples: transmembrane proteins, transcription factors
- If we need to make an annotation only based on specific experimental evidences → we cannot curate anything
    - Not everything can be tested and proved (e.g. transcription factor: the binding is inferred on the sequence analysis, but the actual binding is rarely demonstrated)
    - What is the hypothesis being tested? The authors often have reasons to think it is a signaling molecule (for example) → they show an assay to show it, but they don't do ALL the experiments to show every aspect of a signaling molecule.
    - No paper will prove everything → we will never have all the evidence to make an annotation

- Everyone agrees that it should be fine to annotate based on prior knowledge and some assay
- For MF and CC
    - Sequence information
    - Assay
    - Hypothesis of the author
- The scope of this discussion is: What to do with GO term for which we can never be a proof?
    - Suggestion: annotate to IDA (assay + sequence)
- When we do annotation: we need to look at the context (use of multiple prediction tools), we should try to get the whole picture of the protein.
- In GO-CAM model : we use multiple evidence because we annotate to the "hypothesis": author statement, prior knowledge, sequence, IDA
    - Annotation should not be different when for regular annotation
    - We should annotate to the hypothesis and not individual experiment
- this is very important to have evidence code IDA (EXP) so that the annotations are propagated.

Proposal: Annotate to the conclusion of the hypothesis that the experiment is testing, not the individual experiment which is tested

Presentation to the consortium

1. Annotate to the conclusion of the hypothesis that the experiment is testing
2. For example, 'intergral to plasma membrane' for immunofluorescence localization to membrane if there is also sequence evidence to support this
3. For example, 'RNA polymerase II DNA binding transcription factor activity'
    a. Note - MF refactoring will account for the fact that not all entities described in the literature as TFs bind DNA

ACTION ITEM: [Kimberly (wiki), GO BLOG?]Publicize this decision to users, curators.
ACTION ITEM: [Tool developers with input from Val] Implement a check for terms for which curators should always use a more specific child term.
ACTION ITEM: [Kimberly] Change name of 'spanning component of plasma membrane' to something like 'transmembrane' to align with what is in the literature.
ACTION ITEM: [Kimberly] Solicit other examples from curators for discussion on annotation calls and documentation.

Protein Complexes

Silvain, Alex, Chris, Sandra, David, Barbara, Kimberly, Edith, Shur-Jen, Birgit

1. **Populating GPADs from Complex Portal**

ACTION ITEMS [Birgit, Tony, Alex]:

- Look at GPI as example shown by Alex looked strange
- Newer annotations have full annotations and will be fixed in GPAD
- We cannot export annotations with new ECO codes to GAF
- Tony imports our GPADs into P2GO so GAFs can be generated from there where possible
- Find out what GAF is currently being made

Proposal:

Legacy GO annotations in CP have no PMIDs:

a. need to infer - new GO_ref?
b. If complex has expt evidence (ECO:0000353) then take PMID from IntAct evidence
c. If no experimental evidence for complex (ECO:0005547) then take PMID from review in xrefs

- Evidence code: our use case is effectively TAS (or new ECO:0005547 [biological system reconstruction evidence based on inference from background scientific knowledge used in manual assertion])

ACTION ITEMS [Birgit, Tony, Alex, ?]:

- Decide on new GO_ref for inferred annotation evidences
- Infer legacy annotations and add to our GPAD

We will have new source:CP and new identifiers:CPX-xxxx.y (versioning!)

2. Extending Complex annotations to Gene products

https://github.com/geneontology/go-annotation/issues/1639

Proposal: CC and BP directly inferred from complex annotation to subunit

MF:
We could infer the catalytic unit from the CP and make an annotation to the GP
We looked at some models in Noctua and will do more testing as to how we can solve this.
We will include combinatorial functions in this context.

ACTION ITEMS:
Expand our GPAD to add inferred CC & BP annotations from complexes to subunit GPs [Birgit, Complex Portal devs

Set up working group to decide what we can infer to MF [Birgit, Kimberly, David, Chris, ???]

3. Do we have to hard-code relationships in the ontology?
We need them if we want to have the relationships in the ontology, so YES, continue hard-coding

ACTION ITEM: Birgit, Kimberly, David, Chris, Paul, ???
Working group on at what level we want to include complexes in the ontology (includes 'capable_of' discussion)

4. Where do CP xrefs go?
DH: should go into general DBxref (NOT definition DBxref)
Action Item [Chris]: Perl script to move - done!

Metrics for annotation production

ACTION ITEMS METRIC WG

Training need for curators (new and old)

Many things ! See slides

Usability issues Jacky

Ways to encourage people to contribute:

- Give credit to curators (ORCID)
- Have a easy curation interface
- Templates in Noctua (gene regulation)
- Other items relevant to curators: documentation, start guide/gift basket
- Contact authors to annotate their papers
- Astrid: confidence level / biological context is missing from the GO data; especially for logical modeling and also to enable integrating data from varied resources
- How can we make better use of evidence codes: do they need to be connected to quality ? Perhaps certain terms +Evidence codes combinations can automatically get lower confidence levels ?

Action items

Action items for Monday

ACTION ITEM:New qualifiers
Pascale, Kimberly, Chris
Added to https://github.com/geneontology/go-annotation/issues/1517
Added milestone GOC meeting April 2018
HIGH PRIORITY

- We will implement the new qualifier “acts upstream of or within”, with the more specific children “acts upstream of” and “involved in”
- Groups will have a certain amount of time to include the new qualifier in their GAF file
- If by the deadline, some groups have not provided the qualifier, the general one will be added by GOC database as a post-process on the GAF files
- Documentation/training:
    - Definition of “acts_upstream_of”, “acts_upstream_of_or_within” required
    - There will be training for curators to explain how to use these qualifiers (ideally this will be through the Noctua training)
    - Will need to discuss at annotation calls (curators bring questions to calls)

  Strategies to identify annotations that meant ‘involved in’:

- Make use of PAINT
- Need to make tools available to everyone - can Mary provide the regulates AND process annotation results for all groups?

ACTION ITEM: Relation to use for regulation terms
Kimberly: Document decision that for assertions of type “Gene product involved_in X regulation”, we will use default ‘involved in’
In Noctua, curators will use the ‘regulates’ relations, and in the GPAD output, the annotation will be to ‘involved_in’ ‘regulation of X’. Regulation terms not already in the ontology will be automatically added on a regular (weekly?) basis.
New issue https://github.com/geneontology/noctua/issues/516
(no milestone yet)
PRIORITY NOT SET YET

ACTION ITEM: Multiple qualifiers for annotation of a gene product to a given GO term

Kimberly Create guidelines for using multiple qualifiers

https://github.com/geneontology/go-annotation/issues/1649

ACTION ITEM: Guidelines for using contributes_to'

Kimberly

need to come up with a proposal for the next GO mtg for '**contributes_to**'

https://github.com/geneontology/go-annotation/issues/1650

Added milestone GOC meeting April 2018

Tied with discussion of protein complexes

ACTION ITEM: New CC qualifiers

Kimberly

- Survey groups to see how many want to express BOTH active location and just 'incidental' locations
- Define exactly the relation name: is the relation 'part of' be used for CC? Suggestions: 'localises in'; 'found in'
- Legacy annotations: will default to 'part of', but groups can move to 'enables_activity_in'. We'll also use the same strategies as for BP to assign 'enables_activity_in'
- **DEADLINE**: Proposal: By the next GOC meeting every group will have added the qualifier (have to edit your GAF production pipeline) or a default value will be assigned by GOC (can change them in the future)
- Tell Tony what the qualifiers are to put in P2GO
- Each group needs to update their GAF-generation pipeline to integrate the qualifier. For each group that does not provide any information in the qualifier, these will be automatically filled in by the GOC to the more general qualifier ("part of" for CC/ "causally upstream of or within" for BP)

https://github.com/geneontology/go-annotation/issues/1639

Added milestone GOC meeting April 2018

ACTION ITEM: Proposal to obsolete co-localizes with

Kimberly

- Review EXP annotations using co-localizes with
- Consider creating new GO terms if necessary (for eg, 'peripheral to ribosome')
- Consider creating more precise qualifiers, for eg 'transiently localizes with' or 'dynamically localizes' …

Updated https://github.com/geneontology/go-annotation/issues/1500

No milestone yet

Tied with discussion of protein complexes

ACTION ITEM: Protein complexes representation in Noctua

Birgit & Kimberly Separate working group for protein complex representation in Noctua: Formulate the issue; make a proposal
https://github.com/geneontology/noctua/issues/517
Added milestone GOC meeting April 2018

ACTION ITEM: Getting Noctua ready for production
Chris: List remaining tickets that need to be closed for Noctua to be 'production-ready'; prioritize; define deadline
Where is the list of blocking issues?? Currently there is a single showstopper bug:
https://github.com/geneontology/noctua/labels/bug%20%28A%3A%20showstopper%29

ACTION ITEM: Providing documentation for consuming GPAD files made by Noctua
Chris: document where to slurp up GPADs
https://github.com/geneontology/noctua/issues/518
- No milestone yet

ACTION ITEM: End-user tools for displaying Noctua models
Create a working group to define the needs of end-user tools for displaying Noctua models
David, Chris are interested (Sandra to if this includes a bigger choice of output formats)
PIs Need to prioritize the development of end-user tools for displaying Noctua models
https://github.com/geneontology/noctua/issues/519
No milestone yet

ACTION ITEM: QC for Noctua models
Pascale, Kimberly + QC working group think about how we'll do QC for Noctua models
https://github.com/geneontology/go-annotation/issues/1651
Added milestone GOC meeting April 2018

Action Items Corvallis 2017/06: Noctua functionality

- Provide ways for users to recover and digest GO-CAM units (Gene Ontology-based Causal Activity Model). Ideas include rule-based generations of text statements from model, cytoscape view of network described, etc.
    - (Seth?) I believe that this is a more exploratory and open project. We now offer an RDF endpoint, in addition to the APIs we already have, as well as the proposed feeding from AmiGO. If this is still open in the future, it may be very good Hackathon material. Needs more discussion.
    - https://github.com/geneontology/noctua/issues/519
- ECO codes available for use in Noctua should show how they map up to a classic GO code, and there should be an alert for curators when they are using a code that does NOT map up to a classic code

    - (Seth?) There is no ticket for such functionality. This would likely have to be written into Minerva or as a pre/post check on (some) server.
    - https://github.com/geneontology/minerva/issues/151
- Working group discussion of evidence on complex Noctua models (Kimberly?)
    - https://github.com/geneontology/noctua/issues/517

Action items for Tuesday

ACTION ITEM: Working with external groups

Paul +Ruth Write up recommendations and guidelines for next time we engage an external group of experts

https://github.com/geneontology/go-annotation/issues/1652

Added milestone GOC meeting April 2018

ACTION ITEM Documentation (Proposal to be reviewed by Kimberly/Seth/David)

- Documentation 'landing page'; ie just the list:
    - editors: http. Readthedocs.ontology
    - annotation: (on the wiki; hopefully everyone can find it)
- Page to list github handles of each group. I dont know if this belongs in the 'go-annotation' repo or directly in gene-ontology. We need to know how to reach all data providers
- Start reviewing documentation: suggestion (David, Kimberly)
- Organize a doc-a-thon (December 2017?)
    - Including SOP to produce documentation: document status: "draft", "under review", etc ?
- Develop training documentation (ideally start by collecting training documentation from different groups)
- Document Procedure for using the GitHub repositories (especially go-annotation repo: each concerned db will have a representative assignee; once a group has dealt with issue, remove yourself as assignee; the last assignee closes the ticket
- Priority for documentation: "confidence in the GO" (arising from the HTP presentation)

Added to https://github.com/geneontology/go-annotation/projects/5

ACTION ITEMS: InterPro2GO pipeline

Paul T

- Collect uber 'black list' and post it somewhere that everyone can access it
- Come up with some rule for filtering ahead of time - new GO rule for filtering InterPro2GO for GO pipeline - https://github.com/geneontology/go-site/issues/436
- Put together working group of interested individuals, have a couple meetings, come up with guidelines by May 2018 GOC mtg
- Each of the database should provide their QC pipeline to make sure nothing falls through the cracks

https://github.com/geneontology/go-annotation/issues/1653

Added milestone GOC meeting April 2018

ACTION ITEM: QC rule for ISS: HTP-types will not be allowed as evidence for ISS

https://github.com/geneontology/go-annotation/issues/1654
Added milestone GOC meeting April 2018

- For PAINT : Marc and Pascale decided that HTP are allowed for PAINT; the annotation guidelines is that HTP annotations for components must be supported by MF and/or BP annotations

ACTION ITEM: QC for HTP

- Ask groups to review their papers based on the list the HTP WG has generated https://docs.google.com/spreadsheets/d/11xExGJfj_39xPQUGkam3Xvtd6dtZ5DfANXhM2ZtDYB0/edit#gid=0
    - https://github.com/geneontology/go-annotation/issues/1655
    - Added milestone GOC meeting April 2018

ACTION ITEM: for Transcription factor decision trees

- create transcription regulator activity: https://github.com/geneontology/go-ontology/issues/13588
- DONE PG - still need to merge

ACTION ITEM: PAINT

- Huaiyu Update PAINT documentation: tool instructions + guidelines
    https://github.com/geneontology/go-annotation/issues/1656
- Ahushya/Huaiyu: automatically generate reports to replace the Google docs to track progress /lock families, etc
    - https://github.com/geneontology/go-annotation/issues/1657
- Ahushya/Huaiyu: Challenge mechanism from PAINT: finish the pipeline
    - https://github.com/geneontology/go-annotation/issues/1658

DONE ACTION ITEM: Citation policy

Paul

- Current page: http://www.geneontology.org/page/go-citation-policy
- Page that Paul has been working on: http://www.geneontology.org/page/citing-go

FIXED; updated

GO SLIMS TO DISCUSS/TO DECIDE

- (Suzi): We need to establish **criteria** for creating and maintaining slims - WHETHER we will create/maintain the slims
- We need to establish **guidelines** for creating and maintaining slims
- Which slims are useful? Who is using them?
- Using slims to prioritize areas to curate ?
- Chris MIAAGS: metadata about slims - TO DOCUMENT; probably some decisions need to be made about the exact fields that this will contain
- https://github.com/geneontology/go-ontology/issues/14362

ACTION ITEM: [Kimberly (wiki), GO BLOG?]Publicize the decision of annotation to the hypothesis (including impact on the annotation of tx factors) to users, curators.
https://github.com/geneontology/go-annotation/projects/5

ACTION ITEM: [Tool developers with input from Val] Implement a check for terms for which curators should always use a more specific child term. ("Do not manually annotate"?)
Started: https://github.com/geneontology/go-annotation/issues/1648
https://github.com/geneontology/go-annotation/issues/1659

ACTION ITEM: [Kimberly] Change name of 'spanning component of plasma membrane' to something like 'transmembrane' to align with what is in the literature.

ACTION ITEM: [Kimberly] Solicit other examples from curators for discussion on annotation calls and documentation. ONGOING

ACTION ITEM: COMPLEXES WG Working group on at what level we want to include complexes in the ontology and do we hard-code MF ad BP relationships (includes 'capable_of' discussion); how to represent complexes in GO-CAM models; how to represent MFs of individual complex subunits in GO and GO-CAM models; populate GPAD with legacy CP annotations COVERED IN GITHUB

ACTION ITEMS METRICS WG

ACTION ITEMS TRAINING WG

ACTION ITEMS USABILITY WG

- Working group to discuss additional functionality:
    - Can we capture the Figure from which the data comes from ?
    - Could we capture a confidence score?