GO Consortium Meeting - Bar Harbor, ME - September 26-27, 2003

[Next Meeting: Stanford- SGD organizing: - GO Users Open Mtg; Jan. 15th. GO Consortium Mtg. Jan. 16-17.]

Opening Comments:

Meeting organization: We are a very cohesive group that works well together and we want this to continue. Therefore as we grow in size and in objectives we must continuously address the effectiveness of our organization in order to maintain 1) effective communication, 2) the quality of what the project produces and 3) informalities of the group, so that all feel welcome to contribute and comment. At this point the group has grown to the extent that we must adjust and strengthen the structure and organization of the GO Consortium.. We recognize that there are four major sub-groups here: 1) Ontology Development, including Interest Groups; 2) Annotation; 3) Database and Software Development and 4) Production and Distribution. In this context, we need to discuss how to go about revising the structure of the GO Consortium meetings. For example, the 'whole' group meets less frequently and sub-group meet more frequently. This topic was a thread through the meeting and there was further discussion at the end of the meeting.

1) Group Participant List

EBI-Ontology group	(Midori Harris, Jane Lomax, Jen Clark, Amelia Ireland)
Berkeley DB group	(Suzi Lewis, Chris Mungall)
FlyBase	(Michael Ashburner, Rebecca Foulger)
SGD	(Mike Cherry, Rama Balakrishnan, Maria Costanzo, Rob Nash)
MGI	(Judy Blake, David Hill, Harold Drabkin, Martin Ringwald, Mary Dolan, Li Ni, Joel Richardson, Janan Eppig, Alex Diehl)
TAIR	(Tanya Berardini, Suparna Mundodi)
SWISS-PROT	(Evelyn Camon, Daniel Barrell)
Sanger Parasite Group	(Matt Berriman)
S. pombe/Sanger	(Val Wood)
WormBase	(Eimear Kenney, Kimberly Van Auken)
DictyBase	(Rex Chisholm, Pascale Gaudet, Warren Kibbe, Cathy Li)
GKB	(Lisa Matthews)
RGD	(Susan Bromberg, Norie De la Cruz, Victoria Petri, Mary Simoyama, Lan Zhao)
TIGR	(Michelle Gwinn)
Incyte	(Allan Davis)
ZFIN	(Doug Howe, Sridhar Ramachandran)
Gamene	(Pankaj Jaiswal)

2) Updates on Action Items from St. Croix Meeting

The full listing of Action Items from St. Croix Meeting is at end of report. Most Action Items are completed. Not_Done or In_Progress or Special_Notes items listed here.

1. Update all gp2protein files in CVS. *Need to send reminders to some groups.*
6. BDGP(SwissProt): Request for tool for tentative assignment of GO terms *Not Done*
8. Assemble 'methods' references for IEA. *In progress* - work done by Midori and Michelle. GO is going to maintain a set of generic references of descriptions of IEA techniques for databases to use who themselves do not have reference collections to call on. These will then allow users of the data to distinguish between the different ways that GO terms have been assigned that fall under the IEA umbrella. [Action Item 34, BHmtg]
9. IEA- BDGP to explore means of including larger number of associations in DB and AmiGO. *In progress* [see Action Item 28 BHmtg for a related topic, that is, removing defunct associations].
10. IEA - BDGP to add filtering that is combination of evidence code and reference *Not Done* (needs number 8 to be completed first.)
32. Transcription factor issue…Interest group is going to fix and report. *Not Done*

3) Reports from EBI-GO

3) Reports from Ontology Development Interest Groups
Beyond the reports from the Interest Groups, there was considerable discussion about how to involve more experts in certain biological areas in the development of the ontologies. Lisa reported considerable success for GKB by going to specialty meetings and approaching individuals to discuss GKB and elicit their help. Also, follow-up site visits to researcher's institutions might help. GKB uses a powerpoint template to guide contributors. It was decided that GO should also take this proactive approach [Action Item 3]. While the use of ppt is not applicable to GO it is clear that a comparable user guide and standards are needed for newbie ontology contributors. [Action Item 47]

a) Physiology
Tanya and David provided a file with revisions for physiology section of Process. This will be implemented. Complete revision with terms and definitions available from SGD, MGI, others.

b) Plants Interest Group
We have revamped all the extracellular component terms and are now rearranging and expanding the children of the sexual reproduction terms.

sensu Magnoliophyta

-A problem was discovered with the *'sensu* Magnoliophyta' terms. Many of these terms seem misleading because they actually refer to phenomena that also occur more broadly outside Magnoliophyta. However it was pointed out that that *'sensu* Magnoliophyta' just means 'in the sense of Magnoliophyta' and so does not exclude annotation of non-flowering plant gene products to such a term.
-One alternative would be to replace the word 'Magnoliophyta' with a *sensu* word that could apply equally all groups (that might be annotated with such a term). This would be quite time consuming because we would have to check each annotation case using the term, whether it applied to all plants and whether all green algae were included etc.
-At the moment there are no non-flowering plant species being annotated and so there is not an urgent need for terms to be created for the annotation of non-flowering plants.
-With these points in mind it was decided that we should concentrate on making the flowering plant terms exhaustive and stick to *'sensu* Magnoliophyta'. We will create terms for non-flowering plants when non-flowering plants are being annotated.

4) Reports from Annotation Groups
The following groups submitted progress reports of their activities since the last GO Consortium meeting.
    a) FlyBase - ok
    b) TAIR - ok
    c) MGI - ok
    d) SGD-ok
    e) GOA-ok
    f) WormBase -ok
    g) TIGR - ok
    h) Sanger Pathogen - ok
    i) Incycte - ok
    j) RGD - ok
    k) ZFIN - ok
    l) DictyBase-ok

5) Ontology Development Issues
a) Logical consistency checks
In the documentation there is an example of a logical relationship: If A is a part of B and C is an instance of B, then is A must be a part of C? Then there is an example with "cytoplasm". Jane notes this logic isn't always true in the ontologies and ask if can we fix this?

This lead to a discussion of "part of" and how we use it in GO: Chris (and John) said there are 4 types of part of (letting A represent the 'larger' component and B represent the sub-component)

1. B is sometimes part of A
2. B is necessarily (always) a part of A (this is the one we almost always use)
3. A necessarily has part B
4. A necessarily has part B –and- B is necessarily a part of A (both directions of relationship)

Chris: Technically what many ontologies do is to use the weakest relationship (#1) as the default because it is assumes the least. These relationships can then be adjusted to become more restrictive (and precise) as more is known. In practice, we (GO) already are using the part-of relationship in the stricter sense of #2-- most of the time. (as an aside, Chris met and discussed this with Stuart Aiken in Edinburgh. He is also thinking about this and doing a lot of work in this area). Chris also described the distinction between 'part' and 'proper part'. A proper-part is a direct part and therefore is not transitive. *E.g.*: "a nail is a proper part of a finger and a finger is a proper part of a hand but a nail is not a proper part of a hand". There were several decisions made. First, we agreed to update documentation as it regards the use of 'part-of'. Second, we agreed to henceforth only use 'part-of' in the sense of type #2. Third, we agreed to track down all cases that do not use 'part-of' in the sense of type #2 and restructure the ontology as needed. [Action Item 5, 16]. Fourth, we will consider adding all the different logically distinct 'part-of' relationships because these may prove to be needed in many cases in the future.

b) 'Signal Transducer Activity' term disagreements

Question is whether the current "signal transducer activity" term is appropriate for GO.
Harold/David think it is. They proposed a new definition: "the activity of converting one type of signal into another type of signal" (signals can be light, chemical, etc.) They say the process of signal transduction is more than one step but the function of "signal transducer activity" is the first step.
Amelia: There was an issue with "receptor binding" and "signal transducer activity" - not all signal transducers are receptors. If a receptor is under signal transducer activity it should be involved in signal transduction. If a change is made to the definition of "signal transducer activity" than it should be obsoleted, even though there are lots of annotations to it. Especially since Amelia feels the term has been used incorrectly. Report is attached at the end of Meeting Notes.
Midori: there is the question of whether there is a molecular activity of "signal transducer activity".
Amelia: What about steroid receptors that move steroids in/out of cells?
Many: should we change the wording, add a comment?
RESOLUTION TO "signal transducer activity" question: [Action Items 4]
　　　　　-need to obsolete the current term
　　　　　-make a new term with the same name but a new definition
　　　　　-create the new definition to everyone's satisfaction (to be ironed out later)
　　　　　-add a warning to the comment on the appropriate use of this term
　　　　　-clean up the children terms - some need to be moved to other areas of the ontology.

c) Presence/Absence of function grouping terms

Midori: A couple meetings ago it was decided to remove from Function those terms that grouped things based on something other than activity - like Processes or Components. But, having the grouping terms is useful for annotation so people are in no hurry to remove them. ex. "defense immunity protein activity"
This term is a grouping term found in the function ontology that is solely based on Process and has many children terms where this is also true. We don't want function-terms that represent a process because 1) it is a process, not a function and 2) any is-a relationships of child terms to this parent is illogical. While tempting we don't want terms grouped in Function by nature of being in the same Process.
Judy: Maybe we're trying too hard to put a function on everything and are wanting these function terms when really we should just have a process and no function.
Suzi: Some problems come back to the fact that there is a relationship between function and process which we don't reflect.
Midori: Agreements at meeting don't always manifest into agreement after meetings in email.

Judy: If there is angst, then we need more discussion and to resolve things at meetings. General agreement at meeting can break down in fuzzy specific instances in emails afterwards.
Judy/Midori: practicality verses purity of function ontology
Rex: Perhaps people don't realize there are analogous Process terms to use. Maybe state more clearly in the emails.
[Action Item 6. RESOLUTION TO process **grouping terms in Function**. We will not use Process to group Function terms unless all of the terms being grouped share the same type of function. GO curators will continue to bring these to the attention of the group via email, if agreement is reached quickly - great. If not, it will be resolved at a meeting. Also, in the emails be sure to point out the Process term alternatives to the Function term. Things in Function should have things grouped by function.]

d) Consistency of Parentage (catalysis and binding)

Amelia: catalysis and binding - sometimes an enzyme activity has parents of both the catalysis term and binding term. Mostly there is only the catalysis parent. Which way should it be? Consensus: enzyme activities should have only the catalysis parent. [Action Item 17.Remove all binding parents to enzyme activities where appropriate. Document the fact that binding is not always a parent of enzyme. Binding only when stable binding occurs]

e) Difference between activation of/positive regulation of/induction of/etc

Evelyn: positive regulation does not equal activation
Consensus: some redundancy, can't make synonyms in all cases - need some new definitions and comments. [Action Item 18: curation team will go through and find these and try to resolve them , redefine them as needed and put notes in comments.]

f) Synonyms in ontology files (this was actually discussed after the old action items but seems to belong with this section).

Michael: following the experiment of integrating GO into UMLS it was clear that "synonym" was being used in many ways. Jane has made a synonym file with all of the relationships in the file.
format: GOid/GO term/ synonym type id /synonym . This info should be in the db and in the GO ontology files not just the synonym file. Discussion on whether to stop using inexact synonyms in favor of entry words - answer was no.
5 types of synonyms (one parent, 4 children):
    related (~)
        %exact (=)
        %broader than (<)
        %narrower than (>)
        %other related (!=)
For broader than and narrower than synonyms one must always ask if the synonym should be a GO term.

[ACTION ITEM 8: consensus and resolution: Put the 5 types into the database. Put the 5 types into the flat files. John will need to make DAG edit work with this. Jane needs to write documentation. Chris will add them to the db. A warning of the new file format will go out prior to implementation.]

6) Annotation Issues

a) Need for Annotation Consistency

We discussed the need for greater attention to consistency in annotations. Our users expect the annotations to be based on shared standards so that they can be compared and used in comparative genomics contexts. We agreed that we need to more formally identify a mechanism/team/process to ensure greater annotation consistency. This effort will include the development or employment of tools to evaluate annotations. [Action Items: 48]

b) ISS and sequence dissimilarity.

When two sequence are similar but are missing some key piece of sequence similarity that tells you that your protein can't have the function in question what do you do? [Action Item 24; 29] Add to documentation - use the NOT field for ISS annotation with sequence dissimilarity.

c) Annotating to Complexes:

This was a major and continuing topic at this meeting. Should we assign function to members of a complex when these members either do not engage in the (typically) catalytic activity, or we don't really know the function of the member? This was a very long discussion. There were two separate problems that were discussed simultaneously (see below). Discussion ranged over both of the problems throughout.

Two separate problems:
Problem 1. There is an ontology problem in that when the function ontology has an enzyme activity and with children "regulation of activity" and "catalytic activity" there becomes a true path violation for the regulator in that it's path goes up to the catalytic activity when it does not have that activity.

This could be solved by removing the "enzyme activity" parent from the regulatory subunit. The regulatory subunit would have as parent "enzyme activity regulator".   People feared that this would remove a link between the regulatory subunit and the function it was regulating. Others said that the link would be preserved with the component ontology term assignments.   There were suggestions to rearrange the ontology - but nothing seemed to satisfy the needs. In the end the decision was to remove the "enzyme activity" parent from the regulator term.

[Action Item 10] enzyme activity" terms will no longer have as children their regulatory subunits. The regulatory subunit will have as a parent "enzyme regulator". We recognize that this removes a link in function between the regulator and the enzyme activity. However, we feel this will be covered in the annotation of the gene to the complex in question.

Problem 2.   What to do when annotating the function of a subunit of a complex when that subunit does not have a known activity on its own. Up to now we have been annotating to the potential of a subunit and therefore would annotate the function of the complex as the function of one of the subunits (this is in the documentation). This is not actually correct of course, since the individual subunits do not have the function of the whole complex. But to not do this would lose the relationship of the subunit to the function of the complex to which it contributes. Ideally we would be annotating the functions of complexes and assigning gene products as parts of complexes with those functions, but many databases don't have the ability to do that.
- Some suggested making relationships between GO's Function and Component ontologies.
- Some suggested not linking function to the subunits (if nothing is known about what they individually do) at all.
- Some suggested adding a qualifier in the association file - suggestions: direct/indirect, associated_with, etc.
- Some suggested modifying the association file format to include a way to indicate that gene products A plus B plus C are needed for a particular function.

[Action Item 11] Regarding the annotation of gene-products that are members of a complex:
1. The complex should appear in the component ontology.
2. Gene products that are members of that complex should be annotated to that component terms.
3. The complex itself (the instance of it in your DB) should be annotated to the appropriate function.
4. Gene products that are members of that complex should (if a more precise functional granularity is not known) be annotated to the function of the entire complex, but must have an additional qualifier added. This mandatory qualifier will be placed in the "NOT" column. The string we will use for this qualifier has not yet been finalized, but the candidates that we have discussed are "associated_with", "component_of", and "contributes_to". Whichever string is

decided upon the consequence is that now there will be two allowed values in the NOT column: These are "NOT" and ["associated_with" or "component_of" or "contributes_to"]. If both NOT and qualifier value are needed for the association then they will be separated with a pipe character '|'.

[Action Item 30] In a related topic - Mike will add "complex" as an allowed type of "DB_OBJECT_TYPE" in the gene association file for those groups who are able to store complexes in their dbs and assign terms to them.

d) Validation of Annotation

Up to now there has been no validation within the data sets or between the data sets. Can we use the test set?   We want the consortium to check annotations.

*Michae*l: Need tool that takes association files, gets proteins, clusters them, presents to annotators the GO terms attached to the clusters, then view. Need to flag things that are ok, but come up in the screen so they don't have to be looked at again. Once something is found that needs attention - send message to contributing db to fix it. First time these checks are run it will be a lot to go through but once that's done, should be (hopefully) fairly easy to maintain. Maybe we should have GO school/camp for 2 weeks.

Suzi: 3 things:
1. take existing annotations and check for consistency
2. have a given set of genes annotated by two methods and check for agreement
3. GO camp/school useful for a. resolving discrepancies, b. new people education

It's very important to check consistency between dbs.

Mike: consistency is a goal and sharing , must share nitty gritty of methods to make this work.

Suzi: maybe we should all use the same tools

Mike: that's what GMOD is for.

David: consistency with component and function will be easier than process, process will be different for different species. Will need to choose wisely what defines a shared process.

[Action Item 49]

7) Resource Issues

a) Report from development group on instantiating GO in Prolog

The underlying structure of the ontologies is going to have a big shift into a logic programming language Prolog. This new paradigm will impact the development and storage of the ontologies, but the annotation processes will remain the same and most users won't see a difference. We will continue to provide the GO in various formats.

Chris Mungall gave a report from the working group that met in Bar Harbor prior to the GO meeting. This group included Chris Mungall, Suzi Lewis, David Hill, Harold Drabkin, Joel Richardson, Jim Kadin, and Alex Diehl.

- GO is a mix of 'stem' terms and 'composite' terms. For example: 'oxygen binding' is a composite term of the compound 'oxygen' and the term of the function 'binding'.

- A more complex term is 'positive regulation of smooth muscle contraction" - it can be broken down into its component parts:

the action in the term is "contraction"
"muscle" is the thing being affected by the action
"smooth" is a modifier for "muscle" and so a modifier for the thing being affected
"regulation" is a modifier of the action
"positive" is a modifier of "regulation"

(Aside to this discussion: What would be the term in an anatomy ontology "muscle" or "smooth muscle" - answer: "smooth muscle" would be a child of "muscle")

We might want to think about GO as a language system. GO terms are highly regular in their structure. They lend themselves to formating:

for regulation terms--> QUALIFIER, "regulation of" PROCESS where PROCESS is "contraction" or "biosynthesis", etc. PROCESS can itself have modifiers.

One can deconstruct the GO terms like this and build a grammar. There is a programming language called "Prolog" that breaks down terms into parts/classes. Steps in using this for GO:
1. take all or part of GO and decompose.
2. Maintain this breakdown in GO itself
3. make "oxygen binding" a cross product of compound term (from a compound ontology) and the function "binding".

Now many parallel hierarchies like transport and binding can be maintained more easily.

Question: if we have a way of generating a compound term should we still maintain the compound terms in GO or just have them made as users need them.

Answer and consensus - we should maintain them in GO.

Phenotype Ontology will produce massive cross product from Anatomy ontology and Process.

We will use the build up process for the first time a term is needed and then the term will get an id and be in the ontologies permanently. There will be a user mode for creating specific terms.

Discussion of Chris's talk:

Michael: mapping of component terms in PO to base GO terms

Midori: will this help sorting parent/child relationships for new terms? Chris - it should

Martin: will there be 1 rule or decomposition or several rules? Chris - it will create standard wording

Judy: Will people who do ontology development need to use Prolog? Chris - No, just need to make sure they add rules as necessary.

Rex: so with a new term from many places, will the tool make the term from the many places? Chris - you will put in the term, the tool will suggest optional add-ons or alternative names, and parent terms for you to review.

Rex: will the tool read an anatomy file? Chris/Suzi - yes, it will. A set of developers will work on the core/primitive terms and annotators will work on derived terms.

Michael: primitive will come from anatomical ontology? Chris - yes

Michael: mouse anatomy will have "head development" a compound term, but here the primitive is mouse head not just head and this term will be used for many types of heads - do we want only one GO term?

David: we should have all head types as children of "head development" (children would be "mouse head development", "fly head development", etc.)

Rex: how will the anatomies be used?, import them all and then sort it out.? - not sure of answer to this one.

Prolog demo:

-run deconstruction - get stem terms:
    regulation([regulation, qualifier(Q), regulates: P])

-Grammar: qualifier regulation of process
    [regulation,qualifier:positive,regulates:[contraction, affects:[muscle, qualifier:smooth]]]

-first step: go through all GO and breakdown into stem terms.

-Test parent/child relationships

Chris showed amino acid test -

term "glycine binding" has a parent "amino acid binding" but needs "serine family amino acid binding" as parent term since glycine has parent "serine family amino acid" in the compound ontology. It showed that GO was missing an intermediate term and suggests what to do - either add glycine as a direct child of amino acid in the compound ontology or make a new intermediate term so that "glycine binding" can be a child of "serine family amino acid binding" This tool should solve the interleukin problem from before.

More discussion:

Michael: this new tool is an easier route to maintainability and communication with other ontologies - what are the downsides - for the GO curation/editorial team there will be transitional pain - but not for the users. Once the transition is done will there be other downsides?

Chris: We will need to maintain the other ontologies.

Judy: It depends on stable contributing base of vocabularies, some are not so stable, but it will likely be approximately what it is now.

David: GO curators need to now maintain all of the member ontologies. If "oxygen" doesn't exist in a compound db, who puts it in?

Michael: need a "buy-in" of base ontologies. If we can insure that all of the ontologies we rely on are around this table or under institutional control, than we don't need GO developers to maintain, just need reliable people to maintain them and provide quick turnaround - except maybe compound and protein families.

Judy: What about UniProt and PIR? PIR/PANTHER families are a possibility

Michael: chemical ontology -- EBI has a real chemist - so there will be work on that eventually at EBI. Cell type ontology is fairly mature, all anatomies must be in CVS.

Rex: We need a clearing house to tell people which anatomies their terms should be involved in - define lines of what each ontology encompasses.

Martin: who can write to files at OBO?

Michael: each file will have a person who does the updates?

Martin: but right now who can write to these files?

Suzi: there is a short list of people with write access.

c) Report on DAG_Edit

Suzi gave presentation for John, highlighting new properties (see presentation for details). There was a question on whether DAG Edit can save changes between two versions - GO curation team says you can save the histories - need to check on this. [Action Item 42]. Question to group on when to shift to new DAG Edit which is ready to go. Will organize a testing people and John will visit users during this period [Action Item 43].

d) Report on AmiGO

Chris gave presentation for Brad (see presentation for details). Test of new underlying data structure was done for the GO term correlations/concurrent assignments tool. "Genes who liked that GO term also liked this one." - it worked fine. This was Action Item 16 from St. Croix meeting.

e) Demonstrations

1. Joel Richardson: Viewing annotation vocabulary graphically: using GraphViz. Currently works on mouse data, has plans to make it generic, maybe it should work off the GO db. SGD has similar tool. Will work to putting these out on GMOD.

2. Eimear Kenney: Textpresso This is a tool for mining the full text of publications for relevant sentences.

3. David Hill: Automated paragraph generation from GO annotations. Attempt to develop rules at making a nice text paragraph based on the annotation and GO terms assigned to a protein. Did this because granting agencies and users want text output. David did test with Pax6. He developed simple sentence structure rules that allow the automatic fill in of GO terms and annotation information and production of a text description of what is known about a protein. This text is generated from underlying data, is basically the reverse of the deconstruction described by Chris. Both are necessary for complete usefulness of the GO system.

Ultimately, it is hoped that GO data will be presented to the user with options on viewing - the normal GO term assignment tables, a graphical interface like Joel's, and a text entry like David's describing the sum total of what all of the GO terms and annotations tell us about the protein.

f) Slots = properties

Slots is being accomplished with the Prolog deconstruction stuff Chris presented.

For slots we need to decompose ontologies, additional relationship types, need axiomatic ontologies (elemental, basic terms).Chris will start decomposing terms - needs volunteers to go through them - David/Amelia/ plant person. Should be done with some testing by next meeting. [Action Item 50]

8) Lingering questions from this meeting

1. from TAIR: TAIR has a pathway to term map and SGD has a map to another term in the same tree but at a different level. How should this be handled? We didn't' return to this question.

2. GO Slim: should 3 files be required when sending in a Slim: 1) Go Slim itself, 2. Go term mapping to GO Slim, 3) mapping of genes to GO Slim.

9) Action Items from this meeting

Ontology Development Action Items

1. Create SOPs for checking of ontology integrity
2. Document process for revision of subtrees
3. Create SOP for getting people into interest groups and other interest group activities.
4. RESOLUTION TO "**signal transducer activity**" question:
   i. need to obsolete the current term
   ii. -make a new term with the same name but a new definition
   iii. -create the new definition to everyone's satisfaction (to be ironed out later)
   iv. -add a warning to the comment on the appropriate use of this term
   v. -clean up the children terms - some need to be moved to other areas of the ontology.
5. Document Logic Consistency issues in regards to 'Part-Of' designations. Following documentation, track down instances that are not always 'necessarily part-of', figure out what to do with them (known examples: proteasome and polarisome)
6. RESOLUTION TO process **grouping terms in Function**. We will not use Process to group Function terms unless all of the terms being grouped share the same type of function. GO curators will continue to bring these to the attention of the group via email, if agreement is reached quickly - great. If not, it will be resolved at a meeting. Also, in the emails be sure to point out the Process term alternatives to the Function term. Things in Function should have things grouped by function.
7. Alex will send in SF ticket on 'regulation of survival gene products' under "apoptosis" and GO team will check it out.
8. RESOLUTION on **synonym types**: Put the 5 types into the database. Put the 5 types into the flat files. John will need to make DAG edit work with this. Jane needs to write documentation. Chris will add them to the db. A warning of the new file format will go out prior to implementation.
9. As needed, add English terms as synonyms.
10. RESOLUTION of the **regulator subunit of enzyme activity** as child of activity question: "enzyme activity" terms will no longer have as children their regulatory subunits. The regulatory subunit will have as a parent "enzyme regulator". We recognize that this removes a link in function between the regulator and the enzyme activity. However, we feel this will be covered in the annotation of the gene to the complex in question.
11. RESOLUTION of "**subunit of complex**" annotation issue: It was decided to annotate the gene products of a complex to the complex with component terms. To continue to annotate the individual subunits to the function of the entire complex but with a qualifier in the "NOT" column - the qualifier will be "associated_with". Therefore, there will be two allowed values in the NOT column: "NOT" and "associated_with". If you need to use both values at once separate them with a pipe. [subsequent discussion as to whether 'associated_with' or 'component_of' would be the better tag.

Action Items specifically for the Go Editorial Office in Hinxton.

12. Add two new curators to the web site 'people page'.
13. Commit the new web site with improved index. (jen)
14. Send URL of function ontology documentation round to group for discussion. (done)

15. Document the difference between a parent/grouping term in the function ontology and a single term in the process ontology.
16. Document the 5 different part_of terms and the fact that we mostly use just one of them (necessarily part of).
17. Document the fact that binding is not always a parent of enzyme. Binding is only a parent when stable binding occurs. Remove Binding as parent where appropriate.
18. Standardize use of 'activation', 'induction', 'positive regulation of'. GO curation team will go through and find instances of "positive regulation of"/"activation of"/"induction of" and try to resolve them, redefine them as needed and put notes in comments.
19. Keep an eye out for any standard operating procedure information coming from the Annotators
1. meetings.
20. GO.evidence.html has a bad link. Fix this.
21. Two new tools were demonstrated. Add these to the tools page: Joel Richardson's Annotation
2. Browser and the Textpresso program that Eimear Kenney presented.
22. The folks in the GO office are to the test the new DAG-Edit for a few weeks prior to release.
23. Jane will write documentation on Synonym Types. Need to send a warning of the new file format prior to implementation.
24. Add to documentation - use the NOT field for ISS annotation with sequence dissimilarity.
25. Change the documentation so that ISS can have cardinality >1.

Add documentation that clarifies the section where it tells annotators that if you are unsure of the function/process of your gene to bump up to the next higher term. Add that if that bumping gets you to the root of the ontology you should then use the "unknown" term for that ontologyAction Items for Annotation Groups and for Annotation Oversight

26. Formally identify an Annotation Oversight Team, they will a) access quality, b) set standards c) evaluate the annotations of contributing groups, d) alert those groups to annotations that may need attention.
27. RESOLUTION: If there are IEA sets of associations that have not been updated in one year, they will be removed from the front page and AmiGO if a call to the submitting group doesn't result in an updated file.
28. RESOLUTION: Use the NOT field for ISS annotation with sequence dissimilarity. Everyone keep the ISS consistency (how much similarity is enough for different groups) issue in mind and think about ways to improve it.
29. Mike will add "complex" as an allowed type of "DB_OBJECT_TYPE" in the gene association file for those groups who are able to store complexes in their dbs and assign terms to them.
30. Need a new tool that will check for situations where annotation of GO terms was made (for ex. to a mouse gene) based on terms added to another gene (for ex. from human) with ISS, but where the annotation of the match protein (in ex. human) has since changed. An email would be sent to an annotator to review the annotation for the mouse gene again.
31. Add documentation that clarifies the section where it tells annotators that if you are unsure of the function/process of your gene to bump up to the next higher term. Add that if that bumping gets you to the root of the ontology you should then use the "unknown" term for that ontology (in EBI-GO List too)
32. Everyone should be using a script to check for formatting errors in the association files before submitting them to GO. SGD and others have such scripts to share.
33. Send comments on text sent out by Michelle for HMMs and pairwise matches IEA references. Send any other text for other types of IEA evidence around for comment.
34. Organize the quality control checking for annotations. Make a tool to do the comparisons. Organize the GO school/camp. We all must buy into the concept of annotation consistency.

Action Items for Software and Database Development and Production

35. Send reminders to groups who need to update gp2protein.

36. Mike's group will be establishing a production manager and will hire someone to do the job. This person will work with Brad on AmiGO, Suzi's group on database validations, various Annotation Groups on standards for GO association files.
37. If there are IEA sets of associations that have not been updated in one year, they will be removed from the front page and AmiGO if a call to the submitting group doesn't result in an updated file. (This is the same as Action item #28, but is here because it affects both groups)
38. An AmiGO request: Add a species filter to AmiGO. This could be done either by the using the identity of the contributing database or independent of source database by using taxon id from GenBank available for the related sequence. Note the SF site should be used for this kind of AmiGO request (hence #40 below).
39. Provide a SF ticket for AmiGO improvements and suggestions; provide focus group for AmiGO improvements.
40. Software/db group send Midori db format requirements for the IEA references.
41. There was a question on whether DAG Edit can save changes between two versions - GO curation team says you can save the histories - need to check on this.
42. Organize testing period for new DAG Edit. Approximately 6 weeks of testing. John will visit users of DAG Edit during the testing period.
43. New flat files in new format should have a different file name format. "function.obo, process.obo, component.obo" these files will be terms plus definitions. Feel we still need all three although with new system then can be combined.

Other Action Items

44. Run a test-set on all the GO tools. Generate a test set of genes for tool validation. Get a responsible person to manage the test system. Post results so users can see the kind of analysis/visualizations provided by a Tool. Nobody made a clear commitment to organize this, but many were in favor of it.
45. Pankaj is in contact with a group that wants to translate GO into several European languages, Arabic and Chinese. Pankaj will talk to this group wanting to translate GO and learn the details of their plans. Need precise input as to how the group would deal with update issues.
46. Develop further documentation for ontology development guidelines so that when we get help from outside experts to develop specific branches of the ontology we have a way to introduce them to some of the basic tenets and standards that are needed in order to do this [EBI editorial staff].
47. Mike Cherry (primarily, but not all by himself of course) to propose and up mechanism/team/process to develop 1) manual methods, 2) automated assessment tools, and 3)documentation to ensure greater annotation consistency.
48. Pursue by all possible means methods for improving consistency of annotations: computationally based on sequence; Comparatively, between alternate methods carried out on same gene sets; through training and documentation (camp?) [Suzi, Mike, Michael, and Judy]
49. Chris will start decomposing terms and David/Amelia/ plant person will work with him to help test the results and change the ontologies as needed.

10) Summary
Proposal for future organization:
    -software will be broken into development group and production group
    -the production group will be handled by Stanford group
    -annotation needs quality control oversight - for now Mike is checking into this.

Should we change the way we organize the GO Constorium meeting schedule?
    -all will be the same for the next meeting (in Stanford)
    - Maybe we should have breakout group meetings for the subgroups (GO ontology development, annotation, software) which report back to the big group. However, many people are vested in several of these areas.
    - Maybe we should have breakout groups for the interest groups which report back.
Mike: there will be 1/2 day available for breakouts. They are expecting the meeting to take 2 full days.
Judy: maybe a series of small group meetings followed by the big group.

Suzi: then the big group would only meet 1-2 times a year.

General agreement on this - suggestion to schedule the big meeting following Stanford meeting after the small meetings have been scheduled.

Addendum 1: Report of Action Items from St. Croix meeting.

1. ALL: update gp2protein on central CVS site. still several need to update.

2. Suparna & Amelia: update metacyc mappings (and check that no functions are mapped to) DONE

3. Amelia: change monthly report file names so they'll sort by date. DONE!

4. Amelia: cron job that mails announcement of each new monthly digest to go-friends this is DONE, in the sense that the auto-mailing works, but not done in the sense that the reporting can be improved (Judy did I get this right?)

5. BDGP, JAX: first prototype to be implemented for properties prior to JAX meeting   DONE (Chris reported at Bar Harbor meeting)

6. BDGP (SwissProt?): need to provide a tool for tentative assignment of GO terms.   NOT DONE

7. one row, one term, one reference, one evidence code. DONE!

8. (IEA) Midori: to assemble method references for IEAs…..stuff to discuss at BH mtg

9. (IEA) BDGP to explore means of including larger number of associations in DB and AmiGO.   IEA db tuning…also need expiring date…NOT DONE

10. (IEA) BDGP to add filtering that is a combination of evidence code and reference.   for IEA…and TIGR, add filter for taxonID, query tool for AmiGO NOT DONE

11. (suspect annotations) Midori et al.: Add some things to documentation to describe procedure for error reporting, whether in terms or in associations.   DONE

12. (suspect annotations) GO-central to add links on main web site to report errors in annotation. DONE

13. (suspect annotations) Brad to add button to AmiGO to mail error reports. DONE

14. (suspect annotations) NOT DONE but this will be part of the new annotation oversight system.

15. ALL: review annotation documentation and send in comments to GO-central (Midori to oversee). nothing sent…EBI-GO updating documentation as per BH meeting

16. BRAD: to add term based page. This would show all gene products and the other terms that had been used on each of those terms. A "other customers who used this term, also used these terms". Amazon dot.com approach. *Not done* in production AmiGO, but has been *done* in test of new AmiGO architecture.

17. JOHN: Need DAG-Edit to warn if there are definitions without terms when saving so that the definitions are not lost. DONE in beta version

18. GO central: for all part-of children in the function ontology, change the relationship to is-a and change wording to 'intrinsic regulator' or 'intrinsic catalyst'. DONE

19. Jane: remove 'activity' from 'binding' terms; DONE!

20. Midori & Jane to dredge up what problems were at end of database save testing; send to John. DONE

21. JOHN: Need DAG-Edit and central repository to work more seamlessly...DB or transparent CVS must be implemented. NOT DONE

22. GO-central improve documentation on synonyms. DONE

23. David organizing physiological process interest group. DONE

24. Physiological interest group is to report on progress next time. DONE

25. GO-central delete references to viruses in the definition of extracellular. DONE

26. GO-central move viral component terms back into intracellular. DONE

27. Midori to send examples of regulation to BDGP and Chris et al. to examine how to correctly indicate and model regulation. DONE

28. Eurie: can now proceed to use gene products in terms with the addition of the suffix class and other situations will be handled in the same way. OK

29. GO-central: Update the documentation to reflect the decision on transporters. OK

30. Amelia: Check on the terms in question and make sure they are consistent with the decision regarding transporters (and other bi-directional functions). DONE

31. Michelle: Originally this AI was to send examples of messed up merges to GO-central for resolution. This was done. There are a few "sensu Eukarya" terms with secondary ids that did not have

"sensu Eukarya" in them (Amelia generated a list of about 10). However, it turns out that it is ok that they are that way because, due to the placement of the old terms in the graph (as children of mitochondrial things for example), it is logically implied that they are Eukaryotic and therefore it is fine to make them secondary ids of Eukaryotic specific new terms. The problem for TIGR arose when those terms with mitochondrial parents were used to annotate some bacterial proteins (even though we knew about the path violations for bacteria) because at that point bacterial counterparts did not exist for those terms and they still wanted to capture the information. Therefore, the new Action Item is for TIGR to fix these annotations now that the bacterial counterpart terms are in GO. Thanks to Midori and Amelia for clarification of this. DONE

       32. transcription factor is wrong (mis-defined and mis-annotated). Interest group is going to fix this and report the solution. NOT DONE

       33. All interest groups to provide short (one page more or less) reports for next meeting.

       34. Jennifer: to provide a mock-up of the GO home page using Sanger style links.

Tried this, but didn't work well.

Appendum 2: Signal Transducer Activity report

signal transducer activity : current def
"Mediates the transfer of a signal from the outside to the inside
of a cell [or cellular compartment] by means other than the
introduction of the signal molecule itself into the cell.

The proposed definition of signal transducer is based on the concepts carried by both "signal" and "transducer".I've been looking over the definitions of the individual components of "signal tranducer"; there are two components:

1. detect signal << what is this?
2. change signal into another activity<<<how?

signal: N. an object (an event; detectable physical quantity or impulse) used to convey information;
     V. to communicate

transduce: V.to change one energy or a message into another form (sense organs transduce physical energy (light, heat,) into chemo/electric energy.

transducer: N. a device that converts energy from one form into another.

A signal transducer would thus be a gene product that converts one type of signal into another.

This would distinguish it from the process of signal transduction, of which the transducer would be only one part of the whole process.

A distinction might be made between the type of signal in and out. The transducer needs to change the signal type. However, this may need to be broad: light to chemical is easy, but ligand binding to cause activation of intrinsic enzymatic activity might be harder; both are sort of chemical events although you could consider the ligand binding causing a conformational change as being chemical

enzymatic activity. The enzymatic activity itself is separate.

Objections:
Signal transduction. The biochemical events that conduct the signal of a hormone or growth factor from the cell exterior,through the cell membrane, and into the cytoplasm. This nvolves a number of molecules, including receptors, proteins, and messengers.

>>>does not have to be a molecule; it can be light (see further on).

Are all these proteins therefore signal transducing molecules. Certainly cytokines are accepted as signal transducing moleculeswith the ability to induce signal transduction via receptor binding.

>>No, the transducer is the thing that converts one type of signal to another.Cytokines are the signal, not the transducer.

(paraphrasing) "To me, signal tranducer..." or "I see signal transducers..." - you both have a concept of what a signal transducer is, but I think that the current def and the new def fail to capture it.

I think that the term 'signal transducer activity' has been used to describe the activity of anything involved in a signal transduction cascade, and by using the term thus you are not capturing any more information than you already have by annotating to the process term 'signal transduction'. If you want to have a term to represent conversion of one type of signal information into another, I think it should be a new term because I don't think that 'signal transducer activity' will have been used in this way.

A signal transducer would thus be a gene product that converts one type of signal into another."

it seems possible that more than one of the proteins in a signal transduction pathway could be signal transducers, but not necessarily all of them since they all won't change the signal to another form.

How is a signal transducer thus defined any different from a transporter? ("Enables the directed movement of substances (such as macromolecules, small molecules, ions) into, out of, within or between cells.").

Substance and signal are not the same things. A substance is always a physical entity; a signal is not. Insulin binding it's receptor is a signal, but so is heat, etc. Binding to a receptor does not mean a substance is then transported into the cell.

Reply to the part about transducer vs transporter: with a transporter, a substance goes in one end and out the other. the "signal" can be a substance (like a phermone), but it doesn't have to be. The transducer converts one type of signal to another (a chemical signal (like phermone) to a conformational change , etc.

...in the transducer vs. transporter debate - I understand the difference between transducers and transporters; however, we've got all the receptor activities lumped under 'signal transducer activity' and some receptors work by conveying the signal molecule into the cell.

Then these should not be called transducers, they are transporters.

We can no longer therefore broadly classify all receptor activities as signal transducers - each receptor activity will need to be assessed and recategorized. Are receptors which transport a signal molecule into a cell therefore not signal transducers? Are they involved in signal transduction, though? Or would we say that there has been a change in the signal type, ie. incoming signal is extracellular steroid molecules, and the outgoing signal is intracellular steroid molecules.

In this above cases, there is no transducer.