Phenotype Annotation

- Why is it needed
- Key difficulties

Principle of Compositionality

- create complex descriptions from simpler one
- can be pre-made or can be made at time of use
  Pre-composed vs. post-composed
- formal means of compostion is provided by Genus-differentia
  e.g. plasma membrane of spermatocyte =
  plasma membrane which is part_of a *spermatocyte*

The building blocks of phenotype descriptions: EQ

- Entity (bearer) such as spermatocyte, wing
- Quality (property, attribute)
  - a kind of dependent continuant

Formally, an EQ description defines:

- a **Quality** which *inheres_in* a **bearer entity**

Kinds of entities which can be bearers of [biological] qualities

- Continuants (3D)
  Cell parts (from GO)
  Cells from (Cell ontology)
  Anatomical entities (from FMA, CARO etc)
- Occurants (4D)
  Processes such as cell death

You can also have composite bearers
e.g. "Cell death in eye", which is composed of an occurant and a continuant

Tour of PATO

- top level designed using formal ontology principles
- see slides for hierarchy
- Divisions are also created based on granularity
- Qualities are divided into Monadic and Relational
  Monadic = A quality of a continuant that inheres solely in the bearer and does not require another entity
  Relational = A quality of a continuant that requires another entity apart form its bearer to exist
- Qualities have a implied comparison to the "normal" so temperature-hot, is implied to mean hot as compared to the wild-type
  NOTE: this implied comparison gets tricky for cases where there is no declaration of a "wild type" -- however fuzzy.
- Relational quality [example]
  Sensitivity of an eye to red light
  The quality inheres_in the eye
  With respect to (towards) red light
  Pheno-syntax
  E= eye Q = sensitivity E2 = red light
- Managing "Absence"
  Annotation patterns for absence are under discussion
  e.g. "spermatocyte devoid of asters"
  E = CL:spermatocyte
  Inheres in the spermatocyte
  Q = PATO:lacks_part
  The quality/relation of missing some part or parts
  E2 = GO-CC:aster
  The quality is with respect to the type "aster"
  NOTE: "somebody" said Barry was wrong about not allowing absence :-)

Marry Ann and Bill Bug - the observation of a "phenotype" is also tied to the mechanism (or method) by which it is measured and is not easy to disentangle.
Chris M - it might be hard but is definitely worth trying.

Q - Is the purpose of PATO to express experimental result devoid of the experimental procedure/design?
A - yes

Q - As expression of the result is intimately tied to the experimental procedure then isn't PATO doomed to fail?
A - no clear answer emerged

Pre vs. post composition

Examples of both provided (see presentation)
The pre and post composed phenotype descriptions are equivalent and the pre composed one can be decomposed to the post-composed on IF the terms used in the pre-composed annotation allow for principles of compositionality.

Comparing phenotypes

We want to compare and query both within and across species
CARO is the first project aimed at demonstrating proof of concept of the above

Case study: Defining plant traits with PATO

Case study: Defining plant traits with PATO
http://www.bioontology.org/wiki/index.php/PATO:Pre_vs_Post_Coordinating

Results:
252/784 terms provided with genus-differentia definitions so far
New terms added to PATO
Can be done for Animal phenotypes as well
Done via an automated OBOL run on all 784 terms. 252 were manually validated, the rest skimmed ... found some systematic problem in decomposing them.

Case study: Bacterial phenotypes
Similar to plant study
26 terms added to PATO
Emerging requirement: Need Ontologies for aggregates of organisms

Q - Is there a tool that other phenotype annotation creators can use perform the above for their own phenotype annotations
A - There is an online OBOL, which may or may not work for all organisms. Contact Chris M with your specific issue.

Measurements
Ontologies provide qualitative partitions on the kinds of entities we fid in nature

We may also want to record quantitative information that comes from "measuring" these qualities. The measurement is not the phenotype.

-- lots of discussion on what variables are actually measured [using some units] in order to determine the phenotype and therefore the measurement IS part of the phenotype.

Phenotype exchange formats

Genotyes and phenotypes
- pheno-syntax
- pheno-XML

General purpose
- OWL (using EQ encoding) [or its equivalent in OBO syntax]

GO annotation files
- works with pre-coordinated terms only

OBD-Phenotype
A database for phenotype associations
Tuned for inference, reasoning and graph traversal
Annotations from OMIM, ZFIN and FlyBase
Currently too small dataset to analyze (2 genes)

Phenote: Phenotype Annotation Tool demo

Demo of Phenotype for creating phenotype annotations
Lots of discussion on what constraints Phenote enforces and does not enforce, mostly centered around 'absence' and 'normal'

- requested features (during discussion):
Ability to retrieve phenotype annotations from a database (as in the dictybase version)

Incorporate obo-edit's full featured term-composer

Enable reading of OWL ontology files

Ability to annotate images by defining region of interest (ROI). ROI need not necessarily square

NOTE: by Jan 1st 2008 OBO-edit will read OWL ontology files or John will quit and go work for industry :-)

Linking Animal Models to Human Diseases (using Phenotype annotations)

ZFIN phenotype annotation
Phenotype = entity + quality
Entities come from a lot of different ontologies
Qualities come from PATO

Example of annotating a mutant eye phenotype
E = optic vesicle
Q = apoptotic
T = during (Prim-15)
Tag = abnormal

Showed the (beta) interface for querying the stored phenotype annotations. This will be released on zfin.org very soon.

NCBO project (annotate mutant phenotypes)
Participating in annotating a target set of genes (about 200) with detailed phenotype annotations (along with Flybase and OMIM)

OMIM annotations were done by 3 people (Nicole, Michael and Eric) to allow inter-personnel variation comparison.

Just one mutant (EYE1?) results in hundreds of annotation assertions in EQ format

NOTE: Talk about having a sort of phenotype-BLAST to identify similar phenotypes across organisms that leverages the ontology structure. (both the ontology DAG and inter-ontology mappings such as b/w homologous anatomical entities)

Talking points and discussion

Collaborative tool development

PATO + OBI + NLP + Image based annotation
|----------|
evidence

Tools	Technology stack	Technical readiness	Functionality
ZFIN	Servlet + web	beta	create formal phenotype statements
Dicty	Perl + web	beta	
Phenote	Java + standalone	beta	
Neuroscholar	Java + mysql	full	PDF Markup and metadata + RDBMS store
NLP tools	Perl + Java + web	alpha	
NeuArt	Java + mysql	full	atlas based segmentation viewer with annotation
Mosaic	Java	pre-alpha	describe phenotypes from pharmacogenomic data and literature
Smart Atlas	Java + RDBMS + GIS	full	image annotation 2D
JINX	Java	alpha	image annotation 3D
Biomediator		full	identifying articles to curate
textpresso		full	identifying articles to curate
pubsearch		full	identifying articles to curate

TODO: Get a list of tools that can actually output PATO formalism annotations

What are the roadblocks?

(particularly for groups using pre-composed terms)
- Post-compositionalists [are good]
  - ZFIN
  - FlyBase
- Pre-compositionalists [will be good]
  - MGI
  - Wormbase
  - Dictybase
  - Arabidopsis
  - Agri
  - SGD
- Road blocks to (using PATO) creating and using structured phenotype annotations
  - Gaps in PATO
  - Absence/Gaps in entity ontologies
    - Chemicals
    - Behavior
  - Momentum (legacy phenotype annotations)
  - Tools / documentation
    - For using PATO
    - For managing existing phenotype data
  - Representation / interface for the users for EQ

Lot of discussion on explaining the equivalence between pre and post composed terms and how that would actually work in practice.

SUGGESTION: Distinguish between PATO (the ontology providing terms for qualities) and the Phenotype Annotation Scheme/Language (the syntax and semantics of the formalism using which we describe the phenotype)

TODO: define explicitly the syntax and semantics of the formalism for creating EQ assertions.

TODO: Write a paper describing the needs and issues in creating structured phenotype descriptions, describe the resources (ontologies, tools etc) for that activity and describe the formalism (Phenotype Annotation Language?) for actually "writing" the phenotype descriptions.
[Chris is writing something?]

TODO: Create a tracker to provide a temporary id for a newly suggested term so that the new term can be used in annotation (and then later the temporary id is replaced with the "official" id when it gets assigned)

Relational qualities and when to use them

Q: How do you represent the difference in phenotype at 1 molar vs. 0.1 molar concentration of a drug
A: the phenotype in this case is represented in relation to a "normal". If this normal is different form the usual interpretation of normal then there should be away to describe that in free-text, which can then be converted to a structured representation later.
- e.g. the "normal" hematocrit value of people living at high altitude is very abnormal for people

living at sea level and this "different normal" has to be declared in a rigorous manner.

Q: is this "different normal" part of the phenotype?
Lots of discussion of whether this is a phenotype or that the original experiment based on which normal is defined is just a bad experiment.

Chris M - there might be lots these "different normal" that we haven't seen yet ...

Q: What is the minimal about of information there needs to be declared in order to determine that some phenotype description is "normal" or a "different normal"?

NOTE: point raised by Wormbase that some communities do not consider "normal" as a phenotype.
NOTE: wormbase doesnot use "normal" instead they use not-abnormal to describe a phenotype where nothing was found to be wrong for the features/characteristics assayed.

NOTE: pursue the use of negation in phenotypic description and a mechanism for including an "unaffected" relation - f/u Fabian.

Discussion on Phenotype annotation formalism (aka Pheno-syntax and Pheno-XML)
See http://www.fruitfly.org/~cjm/obd/formats.html
'*compared-to*' tag was added during the discussion and might not defined at the above URL.

Quantitative statements
NOT DISCUSSED

List of required ontologies
PATO (contact George Gkoutos, g.gkoutos@gen.cam.ac.uk)
Use the tracker listed on the PATO wiki - Do NOT SEND HIM EMAILS
Species anatomies (CARO)
Environment
Unit
Chemical (CheBI) including drugs
GO-BP, GO-CC, GO-MF
Cell
OBI (Assays) [Images]
Relations
Spatial Relationships
(adjacent_to, contains etc)
(dorsal, ventral etc)
Disease
Protein Family
Toxicity (adverse reactions)
SO
Taxonomies

Follow-up meetings and online discussions

Sequence variation
Environment
Phylogeny/"Homology"
Spatial
Assays (OBI)