OBD & Phenote: Tools for Storage and Annotation of Biomedical Data using Ontologies

Nicole Washington*, Chris Mungall, ShengQiang Shu, Mark Gibson, John Day-Richter, Sima Misra, Suzanna Lewis
National Center for Biomedical Ontology • Life Sciences Division • Lawrence Berkeley National Labs • Berkeley, CA, USA
*NLWashington@lbl.gov

ABSTRACT

Today's rapidly increasing amount of sequenced genomes allows for more cross-species comparisons of genes and gene-products than ever before. Current automated investigation of genomes is limited to simple queries of sequence and feature data, and any cross-species relationships based on phenotype and functional data must be accomplished by human intervention. The next generation of intelligent bioinformatics tools that will computationally identify new cross-species functional relationships between genes, mutant phenotypes, and human diseases requires resources for storage and access of experimental data that is annotated in a standardized format.

The Open Biomedical Database (OBD) created by the National Center for Biomedical Ontology is a centralized resource for life scientists, clinicians, and bioinformaticists to store and analyze ontology-based annotations linked to primary experimental data. Standardized annotations of experimental data using phenotype and other ontologies will enable researchers to make sophisticated multiple-species queries related to phenotypes and diseases. OBD will accommodate a wide variety of experimental data, and its development is being driven by our collaborating Driving Biological Projects. Current types of data stored in OBD include mutant phenotypes with their associated genotypes in model organisms, disease state information derived from Online Mendelian Inheritance in Man (OMIM), and meta-data associated with clinical trials. The experimental data is stored and queried as triples (entity, attribute, value) described by ontology terms found in the Core 1 Open Biomedical Ontologies (OBO). We have developed a software tool called Phenote to facilitate annotation of phenotype data in stand-alone and web-based applications. Future improvements include integration of the OBD query and browser functions into BioPortal to deliver experimental results in a researcher-friendly format.

PHENOTE : A Phenotype Annotation Tool

Features

* Phenote is a tool for annotation of phenotype data using ontologies
* Users can browse ontolgies for proper term via term completion
* Users can view relationships, synonyms, obsoletes, and definitions of an ontology term
* Phenotype data is input into external databases using Phenote's pluggable data adapter

Uses

* Is currently being develped as a stand-alone and web-based tool
* Easily modified/customized to suit different user needs
* Design is driven by DBP curratror requirements
* A generic tool will be provided for use by other databases

Step 1
SYNCHRONIZE ONTOLOGY DEFINITIONS

Ontology definitions (such as GO, PATO, CL) are retrieved from OBO through BioPortal into the primary databases and Phenote for use in curration.

Step 2
PHENOTYPE ANNOTATION WITH PHENOTE

Currator enters phenotype data associated with a genotype. Each genotype has a phenotype comprised of multiple phenotype characteristics.

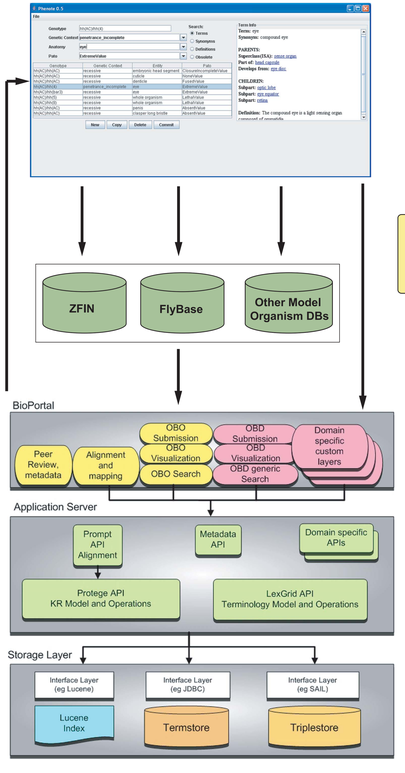
Example: *hh* (FBgn0004644) is_a recessive trait with a phenotype of **incomplete closure** (PATO:0000609), during the process of **embryonic head involution** (GO:0008258).

Step 3
DATA STORAGE IN PRIMARY DATABASES

Data is stored in primary databases such as FlyBase, ZFIN, etc.

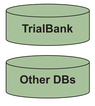
Step 4a
UPLOAD ANNOTATIONS TO OBD

* Formats include phenoXML, OWL, etc.
* Validation steps:
  * formatting
  * versioning



Step 4b
HUMAN DISEASE PHENOTYPES STORED DIRECTLY IN OBD

OMIM disease descriptions will be converted into ontology terms and stored in OBD

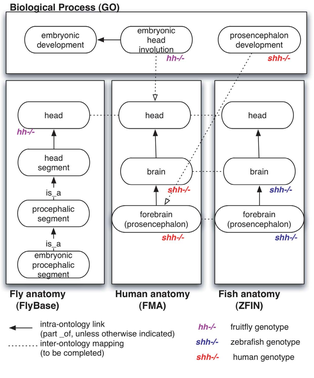


Step 4c
ADDITIONAL ANNOTATED BIOLOGICAL DATA UPLOADED INTO OBD

Other biomedical databases may choose to use their own interface to annotate their data with ontologies and directly upload their data into OBD.

Ontological structure enhances OBD queries

* Mapping between related ontology terms (such as anatomical parts, cell types within an organ, etc.) will link phenotype data based on ontology terms.
* Intelligent query engine can use links to enhance browsing of annotations and enable inter-species comparison



Step 7
COMPUTATIONAL COMPARISON OF PHENOTYPES

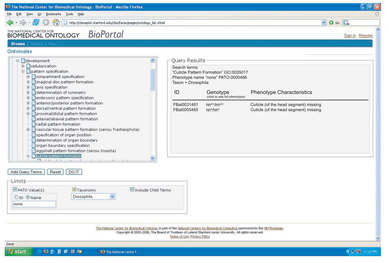
The query (right) started with a phenotype related to the Gene Ontology term "embryonic head involution".

As the diagram (left) shows, this GO term is related to the anatomical part "head", which has anotomical equivalents across the different species' ontologies.

Following the links between the 'is_a' and 'part_of' relationships, as well as additional inter-ontology mappings, we find other mutants that have been annotated with these terms: shh (human) and shh (zebrafish).

Therefore, despite the fact that hh, shh (human), and shh (zebrafish) were annotated with different terms, and have slightly different phenotype characteristics, the phenotypes can be related to one another using ontological relationships in a query.

Access to OBD Phenotype Data through BioPortal*

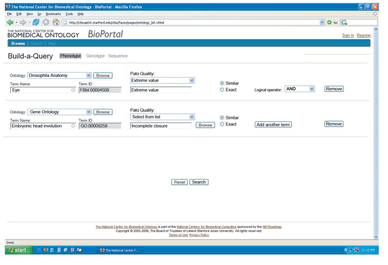


Browser:

* BioPortal will provide a simple tree-browser interface to investigate ontology terms for every ontology in the foundry. More information can be obtained by mousing over terms.
* User can specify search limits. Results display in window frame
* User can re-query their results, find out additional information by clicking on results.
* User can link-out to primary data and literature references

Step 5
EXPLORATION OF OBD DATA THROUGH ONTOLOGY BROWSER

User is interested in any mustants that affect the cuticle. By exploring with the ontology browser, the user finds a relevant ontology term "Cuticle pattern formation", and limits their search to return only those "Drosophila" mutants with a phenotype of "none". After performing the query, several matches are returned to the user. Additional details regarding the mutants, experimental details, and approriate litterature references can be acquired by clicking on the resulting terms.



Build-a-Query:

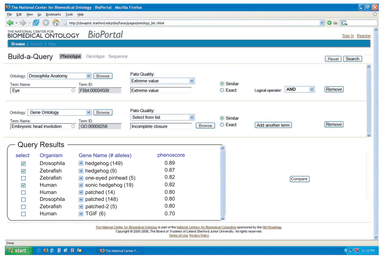
* Three types of queries:
  * Phenotype
    * Ontology term and/or
    * PATO
  * Genotype
  * Sequence
* Query is submitted to OBD and processed for similar phenotypes.

Step 6a
IDENTIFY GENES RELATED BY THEIR PHENOTYPES

User is interested in other genes that give phenotypes similar to the Drosophila hedgehog allele

Because the phenotypes are curated with standard combinations of entity (ontology term)-value (PATO term) pairs, the data can be associated in this way

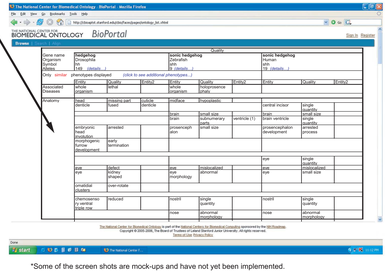
Ex: User enters two phenotypes to match:
an eye defect and
a head developmental defect.



* "Similarity" is reported as the "phenoscore", a computed value based on several factors which may include:
  * # matching phenotypes
  * proximity in "ontological space"
  * sequence similarity
  * organism
  * additional factors
* User chooses some/all of the resulting genes/alleles to compare in greater detail

Step 6b
Query results include several Drosophila, Zebrafish, and Human hedgehog genes, as well as other genes known to be involved in hedgehog signaling (e.g. patched).

Several allelic variant phenotypes are annotated in OBD. The user selects all known hedgehog alleles to compare from Drosophila (*hh*), Zebrafish (*shh*), and Human (*shh*).



Comparison:

* Cross species comparison of common phenotypes
* Different perspectives of ontological relatedness, (levels of granularity) including:
  * Whole organism/disease
  * Anatomical structure/organ
  * Cellular
  * Sub-cellular/molecular
  * Processes
    * signaling networks
    * development
* "Related" terms are grouped together for presentation

Step 8
COMPARE PHENOTYPIC DATA BETWEEN GENES AND ACROSS GENOMES

User selects a set of genes to compare. In this case, the user selects human *shh*, zebrafish *shh*, and fly *hh*.

Multiple alleles are available to describe the phenotypes in more detail.

Relationships (mapping) between ontological terms will allow seemingly different annotations to be related to one-another

*Some of the screen shots are mock-ups and have not yet been implemented.

Funding is provided through NIH grant U54 HG004028.