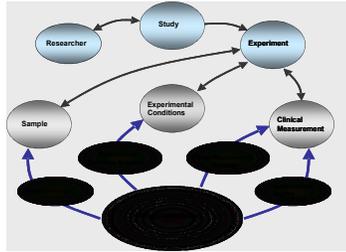


Standardizing Phenotype Data with Ontologies

ABSTRACT: There has been an emergence of multiple large scale phenotype projects in the rat model organism community as well as renewed interest in the ongoing phenotype data generated by thousands of researchers using hundreds of rat strains. In order to provide ready access to such data, the Rat Genome Database (RGD), <http://rgd.mcw.edu> is developing a Rat Phenome Database, <http://rgd.mcw.edu/phenotypes>. To integrate data from multiple datasets and published literature, RGD is developing a Rat Strain Taxonomy (1) and three ontologies to standardize the major aspects of phenotype records: Clinical Measurement Ontology (2), Measurement Method Ontology (3) and Experimental Conditions Ontology (4). In a pilot study phenotype data from Physgen <http://psa.mcw.edu>, the National BioResource Project for the Rat in Japan <http://www.anim.med.kyoto-u.ac.jp/nbr/> and data from published literature were integrated.

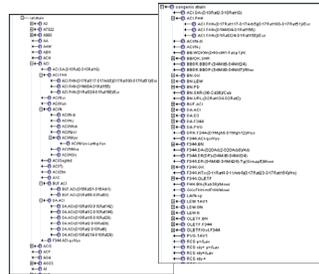
A data mining tool was developed to leverage the utility of the ontologies. (5) Users may begin their search with any of the four components of the phenotype data. In the example, the user chooses a Clinical Measurement (6) followed by a filter for Measurement Method (7) and Experimental Conditions (8) and finally limits the search by Strain (9). The tool keeps track of the search parameters and provides a running count of records retrieved. Only filter choices applicable to set parameters and for which there is data are presented to the user. Results can be downloaded in both summary and expanded formats (10).

Development of the ontologies for use with rat phenotypes continues and expansions and modifications are being applied for use with the human clinical data from the Family Blood Pressure Project in conjunction with Washington University School of Medicine.



Major components of phenotype data records

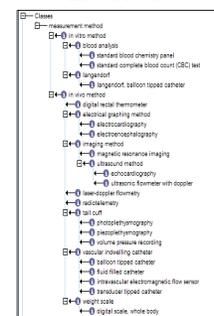
1 Strain Taxonomy



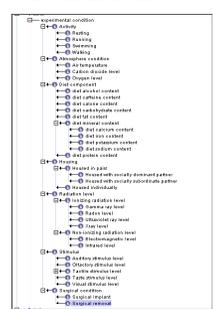
2 Clinical Measurement



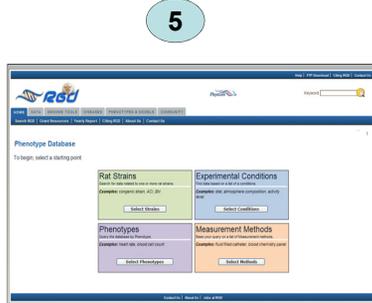
3 Measurement Method



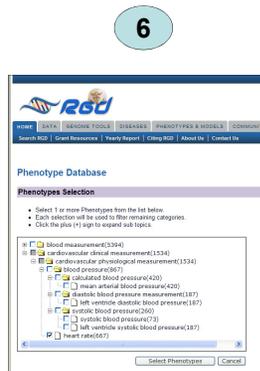
4 Experimental Conditions



Data Mining Using Ontologies



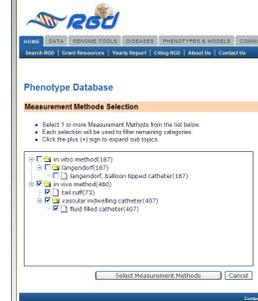
Users may choose to start from any of the 4 phenotype components



The ontology is presented with number of associated records



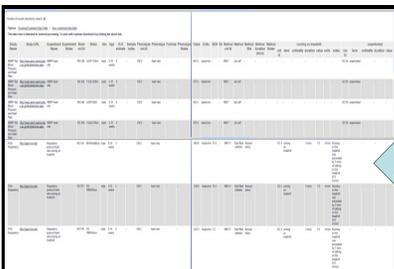
Tool keeps track of query and number of records returned.



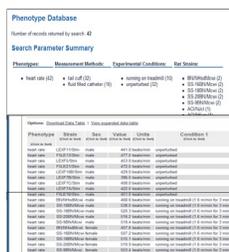
Second filter chosen



Query parameters and record counts shown



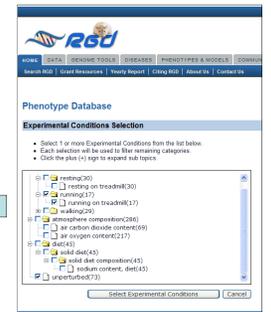
Results may be seen and downloaded in summary or expanded formats



Query record maintained



9



Only applicable choices for which there are data are presented

