NLP and Phenotypes: using Ontologies to link Human Diseases to Animal Models

Washington, N., Gibson, M., Mungall, C.J., Ashburner M., Gkoutos G., Westerfield M., Haendel M., Lewis, S.E.

The path to disease gene discovery in humans is often a lengthy one, but can be significantly shortened if links between human and model organism *phenotypes* are readily available. Collecting and storing these descriptions in a common resource, recorded with ontologies, as well as developing the tools for annotation, access, and analysis are among the goals of the National Center for Biomedical Ontology. The use of well-structured, expert-reviewed ontologies during curation allows biological data to be understandable by both humans and computers, and thereby increases the capacity for meaningful analysis. We have developed the *EQ annotation model*, which uses ontology terms to label and link together *entities*, such as anatomical structures, with the *qualities* describing them. Phenotypes are represented in our model using any combination of entity (such as anatomy) ontologies in combination with an ontology of qualities (PATO). Together with the model organism databases Zfin and FlyBase, we are evaluating this model, using the *Phenote* Annotation Tool to capture the mutant phenotypes of 200 genes known to cause human disease (from OMIM records) that have corresponding fly and zebrafish mutant phenotypes. The phenotypic data modeled in this way is available from the NCBO Open Biomedical Database (OBD), which has the same underlying *annotation* data model, and can currently be accessed via a computational (REST) interface for utilization by other external application or databases. This work is funded by the NIH.