

## THE PHENOGO DATABASE SYSTEM: ANCHORING AND ORGANIZING GENE ONTOLOGY ANNOTATIONS TOGETHER WITH CELL TYPE ONTOLOGY

Sam, Lee<sup>1</sup>; Borlawsky, Tara<sup>1,3</sup>; Li, Jianrong<sup>1,3</sup>; Liu, Yang<sup>1,3</sup>; Friedman, Carol<sup>1</sup>;  
Lussier, Yves A.<sup>1,2,3</sup>

<sup>1</sup>Dept. of Biomedical Informatics, Center for Systems Biology, Columbia University, New York, NY; <sup>2</sup>Department of Medicine, Columbia University, New York, NY; <sup>3</sup>Department of Medicine, Section of Genetic Medicine, The University of Chicago, Chicago, IL

**Keywords:** Gene Ontology, Cell, Context, text mining

The Gene Ontology Annotations (GOA) provide formal relationships for the functions, cellular components and processes of specific genes. For multicellular organisms, these annotations can be refined with the phenotypic context in which a gene is expressed (e.g., cell type, tissue, organ), which is a crucial step for understanding the development and molecular underpinnings of the pathophysiology of diseases. In this poster, we propose a database system (DS), PhenoGO DS, which contains GOA entries enhanced with additional phenotypic context that we have derived using natural language processing techniques and the structures of computational ontologies. More specifically, the PhenoGO DS adds phenotypic contextual information (cell types) to existing associations between the gene products and GO terms specified in GOA. The system also maps the context to identifiers that are associated with different biomedical ontologies, including the Unified Medical Language System (UMLS), Cell Ontology (CO), Mouse Anatomy (MA), National Center for Biotechnology Information (NCBI) taxonomy, GO, and Mammalian Phenotype Ontology (MP). We have developed a query mechanism, available through a simple web-based interface, which allows for the recall of entries based on various levels of phenotypic abstraction across species. Using the structures of the UMLS and CO, the system allows for the retrieval of entire classes of biological products and phenotypes. Users can further expand or restrict their queries using cell descendants and subtypes, as well as UMLS-based phenotypes. The PhenoGO DS comprises over seventeen million entries, including 11 taxa classified into nearly a thousand phenotypes and cell classes. To facilitate access to the database, we developed a web portal that is available at <http://www.phenogo.org>, and allows for the retrieval of data in both human- and machine-friendly data formats.

This study was supported in part by NIH/NLM grants 1K22 LM008308-01(PI:YL) , R01 LM007659(PI:CF, co-PI:YL) and 1U54CA121852-01A1 (National Center for the Multiscale Analysis of Genomic and Cellular Networks :MAGNet PI:Califano, co-inv. Lussier, Friedman).

E-mail: [lussier@uchicago.edu](mailto:lussier@uchicago.edu)