# Ontological analysis of collection improves classification of cardinality phenotypes

Sarah Alghamdi[1], Robert Hoehndorf[1,*]

[1]Computational Bioscience Research Center, Computer, Electrical & Mathematical Sciences and Engineering Division, King Abdullah University of Science and Technology, 4700 KAUST, 23955 Thuwal, Saudi Arabia

**Abstract**

Phenotype ontologies formally characterize and classify phenotypes. We analyze the phenotype classes related to the presence of increased or decreased amount of entities and identify potentially misleading inferences as a consequence of the axioms used to formalize these classes. We propose an ontology design pattern to reformulate these cardinality-related phenotypes and show how these can lead to a clearer classification and higher expressivity. Furthermore, reformulating these phenotype classes in MP and HP improves predictive performance in the task of identifying gene–disease associations through semantic similarity.

**Keywords**

Cardinality phenotypes, Collections and collectives, Ontology design pattern, Semantic similarity

## 1. Introduction

Phenotype data is critical for deciphering the biological mechanisms causing a disease. A formal ontological description of phenotype data can assist in identifying and interpreting these mechanisms. Many ontologies cover the domain of phenotypes for specific organisms, such as the Human Phenotype Ontology (HPO) [1] and the Mammalian Phenotype Ontology (MP) [2]. Most phenotype ontologies define phenotypes using the Entity–Quality (EQ) formalism [3].

Here, we are interested in the phenotype ontology axioms related to an increased or decreased amount of entities present within a body. For example, *decreased T cell number* can be defined using the EQ model as equivalent to `has_part some ('decreased amount' and ('characteristic of' some 'T cell') and ('has modifier' some abnormal))` and *decreased lymphocyte cell number* defined as `has_part some ('decreased amount' and ('characteristic of' some lymphocyte) and ('has modifier' some abnormal))`. Based on these definitions, *decreased T cell number* is inferred to be a subclass of *decreased lymphocyte cell number* . However, depending on the specific meaning of *decreased T cell number* and *decreased lymphocyte cell number* , this may be an unintended inference: if *decreased T cell number* and *decreased lymphocyte cell number* refer to *all* T cells and lymphocyte within a body, then the inference of the subclass axiom is not correct because the decrease or increase of the number of T cells does not imply the decrease/increase of the number of lymphocytes in a body. Similar issues arise when formalizing the absence of T cells, as the absence of T cells does not usually imply the absence of lymphocyte. We identify 2,341 such cases in the MP and 1,119 in the HPO. Among those classes, 490 MP classes and 57 HP classes refer to increased or decreased cell types.

The underlying problem here is that cardinality phenotypes use in their definitions a class that has individual entities as instances whereas single entities (such as a T cell) should not be counted; instead, cardinality is a quality of a collection of entities. We rely on the work of Wood and Galton [4] which previously analyzed collections in formal ontologies.

## 2. Results

We explicitly introduce *collections* of anatomical entities; in particular, we introduce *maximal collections* of entities with respect to another entity, which we define as the collection of all entities of a particular type that are (spatially) contained within another entity. For example, in addition to the class *T cell* which has individual T cells as instances, we introduce a class for the collection of all T cells within a body. Based on collection classes, we can formulate a design pattern that represents the cardinality phenotype, where the entity in the EQ method is replaced with the collection of entity class that we have formulated. Restructuring MP and HPO using our ontology design pattern improves identification of gene–disease association through semantic similarity.

*2022 International Conference of Biomedical Ontologies (ICBO 2022)*

*Corresponding author.

sarah.alghamdi.1@kaust.edu.sa (S. Alghamdi); robert.hoehndorf@kaust.edu.sa (R. Hoehndorf)

0000-0001-5544-7166 (S. Alghamdi); 0000-0001-8149-5890 (R. Hoehndorf)

## References


[1] S. Köhler, M. Gargano, N. Matentzoglu, L. C. Carmody, D. Lewis-Smith, N. A. Vasilevsky, D. Danis, G. Balagura, G. Baynam, A. M. Brower, et al., The human phenotype ontology in 2021, Nucleic acids research 49 (2021) D1207–D1217.

[2] C. L. Smith, J. T. Eppig, The mammalian phenotype ontology as a unifying standard for experimental and high-throughput phenotyping data, Mammalian Genome 23 (2012) 653–668. URL: https://doi.org/10.1007/s00335-012-9421-3. doi:10.1007/s00335-012-9421-3.

[3] G. V. Gkoutos, P. N. Schofield, R. Hoehndorf, The anatomy of phenotype ontologies: principles, properties and applications, Briefings in Bioinformatics 19 (2018) 1008–1021.

[4] Z. Wood, A. Galton, A new classification of collectives., in: FOIS, 2008, pp. 109–120.

[5] M. Ringwald, J. E. Richardson, R. M. Baldarelli, J. A. Blake, J. A. Kadin, C. Smith, C. J. Bult, Mouse genome informatics (MGI): latest news from MGD and GXD, Mammalian Genome (2021). URL: https://doi.org/10.1007/s00335-021-09921-0. doi:10.1007/s00335-021-09921-0.