Mondo Disease Ontology: harmonizing disease concepts across the world

Nicole Vasilevsky[a1], Shahim Essaid[a], Nico Matentzoglu[b], Nomi L. Harris[c], Melissa Haendel[a,d], Peter Robinson[e], Christopher J. Mungall[c]

[a]*Oregon Health & Science University, Portland, OR, USA*
[b]*European Bioinformatics Institute, Hinxton, UK*
[c]*Lawrence Berkeley National Laboratory, Berkeley, CA, USA*
[d]*Oregon State University, Corvallis, OR, USA*
[e]*The Jackson Laboratory, Farmington, CT, USA*

[1]*Corresponding Author: Nicole Vasilevsky, Email: vasilevs@ohsu.edu*

Keywords. Disease, Rare disease, Ontology, Clinical Terminology

1. Introduction

Many terminologies and ontologies currently exist for classifying and describing diseases, each of which has been designed for a particular purpose, and as such has different strengths. Examples include the National Cancer Institute Thesaurus (NCIt), the Online Mendelian Inheritance in Man (OMIM), SNOMED CT, ICD, ICD-O, OncoTree, MedGen, Disease Ontology, and numerous others. Our census of disease ontologies identified 37 such resources. [1] However, these standards partially overlap and often conflict, making it difficult to align knowledge sources. This need to integrate information has resulted in a proliferation of mappings and cross-references between disease entries in different resources; these mappings lack completeness, accuracy, and precision, and are often inconsistent between resources.

In order to computationally leverage the wide array of available knowledge sources for diagnostics and to reveal underlying mechanisms of diseases, we need to understand which terms are truly equivalent across different resources. This will allow integration of associated information, such as treatments, genetics, phenotypes, etc. We therefore created the Mondo Disease Ontology to provide a logic-based structure for unifying multiple disease resources. The goal of Mondo is to integrate the classifications and relationships of commonly used disease ontologies into a single semantically coherent resource to enable the aggregation and analysis of disparate clinical data repositories and facilitate the discovery of relationships between disease concepts across ontologies

2. Methods

Mondo was created by a combination of algorithmic equivalency determination using the k-BOOM algorithm [2], and expert curation. Mondo spans both rare and 'common' disease, comprising monogenic and polygenic (complex, common) diseases, infectious diseases, trauma, and cancer. It provides equivalence mappings to other disease resources, but in contrast to other mapping sets, Mondo precisely annotates each mapping using strict semantics, so that we know when two diseases are precisely

equivalent or merely closely related - allowing computational integration of associated data.

Mondo provides a hierarchical structure which can be used to annotate data at different levels of precision. Terms are classified under multiple parents, as needed, and external ontologies are imported for use in equivalence and subclass of axioms.

3. Results and Discussion

Mondo currently contains over 20,000 disease classes. It is iteratively developed with the community and is under continuous revision, with future plans to further revise the top-level classes. Mondo is an OBO Foundry ontology and can be viewed on the OBO website (obofoundry.org/ontology/mondo.html), or with other ontology browsers such as the Ontology Lookup Service (OLS) (ebi.ac.uk/ols/ontologies/mondo). The ontology files and current releases are available on GitHub (https://github.com/monarch-initiative/mondo), with releases made on a monthly basis. We invite the community to contribute to Mondo; visit our website, GitHub repository and/or join our mailing list (mondo-users@googlegroups.com).

Mondo is being utilized in diverse applications and resources such as the Monarch Initiative [3], ClinGen [4], and Gabriella Miller Kids First Data Resource [5], which is a curated database of clinical and genetic sequence data from pediatric patients with structural abnormalities or childhood cancers. The Experimental Factor Ontology (EFO) employs Mondo's classifications and axioms to facilitate annotation of disease information in applications that use EFO. By integrating information from multiple disease resources, amassed through years of work by researchers, clinicians, ontologists, and other scientists from around the world, Mondo aims to make this knowledge readily available to the scientific community and grow its value through logical connections across resources.

References

[1] Haendel MA, McMurry JA, Relevo R, Mungall CJ, Robinson PN, Chute CG. A Census of Disease Ontologies. Annu Rev Biomed Data Sci. 2018;1: 305–331.

[2] Mungall CJ, Koehler S, Robinson P, Holmes I, Haendel M. k-BOOM: A Bayesian approach to ontology structure inference, with applications in disease ontology construction. bioRxiv. 2019. p. 048843. doi:10.1101/048843

[3] Shefchek KA, Harris NL, Gargano M, Matentzoglu N, Unni D, Brush M, et al. The Monarch Initiative in 2019: an integrative data and analytic platform connecting phenotypes to genotypes across species. Nucleic Acids Res. 2020;48: D704–D715.

[4] Clinical Genome Resource. Welcome to ClinGen. Available: https://clinicalgenome.org/

[5] Working Together to Put Kids First. Available: https://kidsfirstdrc.org/