Integrating and Maintaining OMIM Cross-references in the Disease Ontology

Susan M Bello[a], Cynthia L Smith[a], Lynn M Schriml[b]

[a] Mouse Genome Informatics, the Jackson Laboratory, Bar Harbor, ME, USA,
[a] University of Maryland School of Medicine, Institute for Genome Sciences, Baltimore, MD, USA

Abstract

Integrating information from multiple disease resources into the Disease Ontology (DO) requires understanding how different resources represent disease related information. Here we present how the DO team integrates OMIM into the DO including approaches prioritizing entries and developing guidelines for consistent integration of different types of disease related entries in OMIM.

Keywords:

Disease, Ontology, Integration.

Introduction

As our understanding of human disease changes and evolves over time it is necessary to review and revise disease entity relationships between different disease resources. In the Disease Ontology (DO, http://www.disease-ontology.org (1)) terms from many different disease resources are integrated as database cross-references (dbxrefs) to DO terms. Given the importance of the of the Online Mendelian Inheritance in Man (OMIM, omim.org (2)) as the gold standard reference for human disease and the relation of these disease terms to human genes, integrating and maintaining the OMIM cross-references in the DO is a high priority set of tasks.

Methods

The tasks involved in integrating OMIM into the DO include: refining and maintaining the relations between OMIM phenotypic series and DO terms; identifying cases where changes in OMIM result in a change in the disease attached to a specific OMIM ID; removing obsolete OMIM IDs from the DO; and adding new OMIM IDs to the DO as OMIM adds additional phenotypes. The DO GitHub site currently has 52 reported issues that relate to OMIM disease terms, marked with the label 'OMIM'. These include both ongoing curation tasks and requests for the incorporation of specific OMIM terms from external users. A number of the external user requests for additions or refinement of OMIM phenotypic series representation have been combined to identify those OMIM terms that were requested by multiple groups.

Results

In addition to integrating OMIM entries representing disease phenotypes, our team is developing a set of guidelines for determining whether to incorporate all of the members of the OMIM phenotypic series (representing the genetic heterogeneity of similar disease phenotypes across the genome and including susceptibility and modifier loci). For example, the OMIM phenotypic series 'Parkinson Disease' (PS168600) has 31 entries representing 24 unique OMIM records. Fifteen of these IDs have been incorporated into the DO as cross references to DO terms. For instance, the DO term 'autosomal recessive juvenile Parkinson's disease 2' (DOID:0060368) has a cross-reference to the OMIM term 'PARKINSON DISEASE 2, AUTOSOMAL RECESSIVE JUVENILE' (OMIM:600116). Here the OMIM term represents a subtype of Parkinson's disease defined by mutation in the PARK2 gene. A further 4 OMIM records (OMIM:607688, OMIM:610297, OMIM:613643, OMIM:614251) are related to the DO term Parkinson's disease (DOID:14330) not as cross-references but instead using the relation 'contributes to condition' (RO:0003304). These OMIM records do not represent subtypes of Parkinson's disease but are instead representing susceptibility or risk loci. The remaining 5 OMIM records are in the review process. Some of these (OMIM:613164, OMIM:300557) have titles in OMIM that resemble those of the disease subtypes, however, review of the OMIM text and associated references suggests that these may be susceptibility or modifier loci. Consultation with OMIM regarding these subtypes identified that insufficient evidence was available to decide how to classify these entries at the current time. Thus it was decided to not incorporate these in the DO until further knowledge is presented.

Conclusions

This work has made clear that reliable and accurate incorporation of entities from disease resources into the DO requires both a thorough understanding of disease concepts, coordination and co-development of guidelines and insight into the nuances of representation of these concepts by various resources. Only through careful review and consultation with these resources can we provide an integrated view of disease in the DO.

Acknowledgements

National Institutes of Health–National Human Genome Research Institute (NHGRI) [U41 HG008735-01A1]

Address for correspondence

Susan M. Bello; email: susan.bello@jax.org.

References

1. Schriml LM, Mitraka E, Munro J, Tauber B, Schor M, Nickle L, et al. Human Disease Ontology 2018 update: classification, content and workflow expansion. Nucleic Acids Res. 2019 Jan 8;47(D1):D955–62.
2. Amberger JS, Hamosh A. Searching Online Mendelian Inheritance in Man (OMIM): A Knowledgebase of Human Genes and Genetic Phenotypes. Curr Protoc Bioinforma. Hoboken, NJ, USA: John Wiley & Sons, Inc.; 2017 Jun 27;58:1.2.1–1.2.12.