Expanding the Mammalian Phenotype Ontology to Meet the Needs of COVID-19 Model Curation

Susan M. Bello [1], Anna V. Anagnostopoulos [1] and Cynthia L. Smith [1]

[1] *The Jackson Laboratory, 600 Main St., Bar Harbor, ME, USA*

Abstract

Through the course of the COVID-19 pandemic a wide array of signs and symptoms displayed by patients have been identified. Mouse models of COVID-19 display phenotypes that correlate to many of these signs and symptoms. To capture the phenotypes of these mouse models in the Mouse Genome Informatics database the Mammalian Phenotype (MP) ontology was reviewed to map these symptoms to existing MP terms and add new terms where needed. This review identified over 350 COVID-19 signs and symptoms and resulted in the addition of 127 new MP terms.

Keywords

COVID-19, phenotype, mouse models

1. Introduction

To meet the needs of researchers working to fight the COVID-19 pandemic, Mouse Genome Informatics (MGI, www.informatics.jax.org)[1] incorporated data from mouse models of COVID-19 into the existing knowledgebases. Critical to this integration was ensuring that the Mammalian Phenotype (MP) ontology[2], used by MGI to annotate phenotypes displayed by mouse models, had the necessary terms to cover the range of signs and symptoms potentially displayed by COVID-19 patients and models. As COVID-19 is a newly emerged disease, the full spectrum of signs and symptoms has not been fully defined. To identify appropriate terms, the emerging body of COVID-19 literature and resources were reviewed, and relevant terms extracted.

2. Identification of COVID-19 Signs and Symptoms

The initial list of COVID-19 signs and symptoms was seeded from the list developed in the COVID-19 Virtual Biohackathon 2020[3]. This list was then augmented and expanded by multiple literature reviews conducted using both PubMed and the bioRxiv/medRxiv COVID-19 collection [https://connect.biorxiv.org/relate/content/181] to identify relevant articles. Papers were collected throughout 2020-2021 with resources being accessed weekly or biweekly to collect newly added articles. Emphasis was placed on papers that collected signs and symptoms for larger sets of patients as opposed to reports on individual patients. In addition, COVID-19 symptom and sign lists from the World Health Organization[4] (WHO) and the US Centers for Disease Control[5] (CDC) were incorporated into the list. Symptoms and signs were extracted from over 220 references. The full set of references can be found on the working spreadsheet linked in the Mammalian Phenotype GitHub (https://github.com/mgijax/mammalian-phenotype-ontology/issues) issue #3406.

PubMed was systematically interrogated for peer-reviewed articles with full text availability, using the search terms "COVID-19", "SARS-CoV-2" or "2019-nCoV" in conjunction with "symptoms", "signs", "characteristics", "features", "manifestations'" or "complications". For organ/system-specific features additional search terms were combined, as appropriate. For example, to search for articles

reporting on the ocular manifestations of COVID-19, search terms included "ophthalmological", "ocular", "eye" or "vision".

The bioRxiv/medRxiv COVID-19 collection was manually reviewed to identify non-peer-reviewed preprints. Papers mentioning in the title specific symptoms or systemic effects of COVID-19 in patients were selected for further review. As the pandemic progressed, focus shifted to collecting in depth characterizations of symptoms and reports of symptoms in anatomical systems without existing symptom reports. In addition, with the identification of long COVID[6] and multisystem inflammatory syndrome in children (MIS-C)[7] signs and symptoms for these conditions were incorporated into the extracted signs and symptoms list.

2.1.   Mapping of COVID-19 Signs and Symptoms to the MP

Each COVID-19, long COVID, and/or MIS-C sign or symptom was listed on a Google spreadsheet. The MP and Human Phenotype ontology (HPO)[8] were then searched for a matching term and any identified term in the ontology was added to the spreadsheet. If no match was identified in the ontology, then the symptom was reviewed to determine if it was within the scope of the ontology. For missing terms determined to be within scope of an ontology GitHub issues were created on the relevant repository to request that new terms be added to the ontology. In the MP GitHub issue tracker tickets relevant to COVID-19 have a "COVID-19" tag attached to facilitate tracking of relevant requests.

How closely an ontology term matched a given sign or symptom was indicated using Simple Knowledge Organization System (SKOS) terms[9]. Ontology terms that have the same meaning as a sign or symptom are marked as exact matches. Ontology terms that cover a wider range of phenotypes than the sign or symptom are marked as broad matches. For example, while *Fever* (HP:0001945), defined in the HPO as "Elevated body temperature due to failed thermoregulation.", is covered by *increased body temperature* (MP:0005533) there are other causes of increased body temperature besides loss of thermoregulation that may also be encompassed under the MP term. Of the 204 MP terms mapped to COVID-19 signs or symptoms where the SKOS mapping relation has been assigned, there are 162 exact, 16 broad, 8 close, 15 related, and 3 narrow mappings. SKOS mappings were not recorded from the start of the project so there remain 87 MP terms where an SKOS term is unassigned.

Table 1 lists an excerpt from the full set of symptoms. The full list of terms can be accessed online upon request. The MP GitHub issue #3406 has an Excel spreadsheet with the mappings of MP terms as of June 7, 2021. In the online list terms are grouped by affected system and the common and disease severity classifications provided by the CDC or WHO are marked. Throughout the project the full set of spreadsheets were shared with the HPO team to facilitate coordination of term mapping and additions. Terms present in both ontologies are being mapped to each other and the resulting mappings will be deposited in the Mouse-Human Ontology Mapping Initiative repository (https://github.com/mapping-commons/mh_mapping_initiative).

2.2.   Addition of New Terms to the MP

COVID-19 and related syndrome signs and symptoms that did not match an existing MP term were evaluated for addition to the MP ontology. The most common type of term addition related to signs associated with standard blood work for patients. For example, *elevated blood C-X-C motif chemokine ligand 10 level* was reported in COVID-19 patients (see Table 1). This term was added to the MP ontology as *increased circulating CXCL10 level* (MP:0031220) following the standard pattern for abnormalities of circulating protein level terms in the MP ontology. In addition, terms for *abnormal* and *decreased circulating CXCL10 level* (MP:0031218 and MP:0031219) were added. While these terms do not directly match the COVID-19 sign they allow for annotation of mouse models where genetic or environmental interventions alter the phenotypes exhibited by the mouse model. This new term was then used to annotate the B6.Cg-Tg(K18-ACE2)2Prlmn/J (MGI:6389236) model of COVID-19 based on phenotypes described by Yinda CK et al[10].

Table 1

Example of mapping of reported COVID-19 symptoms to MP and HPO terms

Reported symptom	MP term (ID)	SKOS Match	HPO term (ID)	SKOS match	Reference
Dyspnea	respiratory distress (MP:0001954)	exact	Respiratory distress (HP:0002098)	exact	PMID:32586739
Anosmia	anosmia (MP:0004512)	exact	Anosmia (HP:0000458)	exact	PMID:32464367, PMID:32383370
Fever	increased body temperature (MP:0005533)	broad	Fever (HP:0001945)	exact	PMID:32574165
Interstitial pneumonia	interstitial pneumonia (MP:0001862)	exact	Interstitial pneumonitis (HP:0006515)	exact	PMID:32526193
Hemiparesis	Hemiparesis (MP:0031201)	exact	Hemiparesis (HP:0001269)	exact	PMID:32354768, PMID:32436105
Cephalgia (headache)	No term		Headache (HP:0002315)	exact	PMID:32464367, PMID:32574165
Elevated blood C-X-C motif chemokine ligand 10 level	increased circulating CXCL10 level (MP:0031220)	exact	No term		PMID:32360286, PMID:31986264

3. Conclusion

A total of 127 new COVID-19 sign and symptom related MP terms have been added as of May 19, 2021. These terms have been used in over 400 genotypes involving 360 mouse markers (data pulled from MouseMine.org[11] on June 8, 2021 using the list of new COVID-19 related MP identifiers). These terms were used to identify COVID-19 models and were also used in annotation of discrete mouse models allowing for integration of COVID-19 data with the full corpus of MGI phenotype data. The ability to find mutations that alter the expression of COVID-19 phenotypes independent of infection can then be used to identify potential targets for new treatment strategies.

MGI biocurators continue to expand and refine the MP to reflect the evolving understanding of the COVID-19 clinical spectrum and ensure robust annotation and retrieval of COVID-19 relevant model phenotypes and genes.

4. Acknowledgements

5. References

[1] J. A. Blake et al., Mouse Genome Database (MGD): Knowledgebase for mouse–human comparative biology, Nucleic Acids Res., vol. 49, no. D1, pp. D981–D987, Jan. 2021.
[2] C. L. Smith and J. T. Eppig, The Mammalian Phenotype Ontology as a unifying standard for experimental and high-throughput phenotyping data, Mamm. Genome, vol. 23, no. 9–10, pp. 653–

668, Sep. 2012.

[3] COVID19 Virtual BioHackathon 2020, 2020. [Online]. Available: https://github.com/virtual-biohackathons/covid-19-bh20/wiki. [Accessed: 16-Apr-2020].

[4] WHO Headquarters, Clinical management Clinical management Living guidance COVID-19," World Health Organization, 2021. [Online]. Available: https://www.who.int/publications/i/item/WHO-2019-nCoV-clinical-2021-1. [Accessed: 25-Jan-2021].

[5] CDC, "Symptoms of COVID-19." [Online]. Available: https://www.cdc.gov/coronavirus/2019-ncov/symptoms-testing/symptoms.html. [Accessed: 22-Feb-2021].

[6] Long COVID: let patients help define long-lasting COVID symptoms, Nature, vol. 586, no. 7828, pp. 170–170, Oct. 2020.

[7] N. Nakra, D. Blumberg, A. Herrera-Guerra, and S. Lakshminrusimha, Multi-System Inflammatory Syndrome in Children (MIS-C) Following SARS-CoV-2 Infection: Review of Clinical Presentation, Hypothetical Pathogenesis, and Proposed Management, Children, vol. 7, no. 7, p. 69, Jul. 2020.

[8] S. Köhler et al., Expansion of the Human Phenotype Ontology (HPO) knowledge base and resources, Nucleic Acids Res., vol. 47, no. D1, 2019.

[9] Mapping Properties. [Online]. Available: https://www.w3.org/TR/skos-reference/#mapping. [Accessed: 09-Jun-2021].

[10] C. K. Yinda et al., K18-hACE2 mice develop respiratory disease resembling severe COVID-19, PLOS Pathog., vol. 17, no. 1, p. e1009195, Jan. 2021.

[11] H. Motenko, S. B. Neuhauser, M. O'Keefe, and J. E. Richardson, MouseMine: a new data warehouse for MGI, Mamm. Genome, vol. 26, no. 7–8, pp. 325–330, Aug. 2015.