Tracking the functional effects of SARS-CoV-2 genomic variants: An ontology-driven approach

Madeline Iseminger[1,2,*], Muhammad Zohaib Anwar[2], Rhiannon Cameron[2], Damion Dooley[2], Paul Gordon[3], Emma Griffiths[2], Anoosha Sehar[2], Khushi Vora[3] and William Hsiao[1,2]

[1]*University of British Columbia, Vancouver, BC, Canada*

[2]*Centre for Infectious Disease Genomics and One Health, Faculty of Health Sciences, Simon Fraser University, Burnaby, BC, Canada*

[3]*Centre for Health Genomics and Informatics, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada*

Abstract

Emerging SARS-CoV-2 genomic variants can impact disease transmission, viral antigenicity, infection severity, and vaccine efficacy. As such, it is critical that new variants and their potential impacts are tracked in a rapid and globally accessible way. Due to the intensive labor required to manually extract genomic variant information from the literature, a semi-automated approach is needed. We present a novel ontological framework for describing SARS-CoV-2 mutations and their purported functional effects, and contextual data for literature evidence. This framework follows Basic Formal Ontology guidelines and is interoperable with existing OBOFoundry ontologies. When coupled with genomic surveillance of circulating pathogens, it will assist with rapid sharing of potential functional impacts of new variants in a standardized, machine-readable way. In future, the model could be extended to use cases beyond SARS-CoV-2, such as influenza or antimicrobial resistance. The framework consists of three linked minimodels: variant calling, host and pathogen phenotypes, and literature evidence. The variant calling model describes the process from sequencing a viral sample to variant calling, and linking variant calls to phenotypes. As far as we know, this is the first model in OBOFoundry to describe mutation-phenotype relations. The mutation names exist on the instance level to avoid proliferation of new classes, and they are correlated with punned instances of phenotypes. Sequence Ontology[1] terms for mutation types were not used to remain compatible with BFO standards. The phenotype model contains terms for functional impacts that are correlated with SARS-CoV-2 mutations, spanning levels of granularity from molecular impacts to impacts on disease transmission. The terms are based on terms from Pokay[2], a hand-curated repository of SARS-CoV-2 mutations and their functional effects, with links to related research articles. The phenotype terms are housed in the Pathogen Host Interaction Phenotype Ontology (PHIPO)[3], while non-phenotype terms (relating to vaccines, treatment, diagnostics, and associations with pre-existing conditions or homoplasy) are reused from the Vaccine Ontology (VO)[4], Coronavirus Infectious Disease Ontology (CIDO)[5] wherever possible. Phenotype terms begin with "altered", matching PHIPO[3], as a description of the change taking place. The literature evidence mini model links mutation calls and phenotypes to their literature evidence sources. Short free text descriptions of the research findings are included here. We are in the process of developing a text mining module utilizing the minimodels to explore semi-automatic retrieval of relevant literature.

Keywords

SARS-CoV-2, application ontology, mutations and phenotypes, literature retrieval

References

[1] K. Eilbeck, S. E. Lewis, C. J. Mungall, M. Yandell, L. Stein, R. Durbin, M. Ashburner, The sequence ontology: a tool for the unification of genome annotations, Genome Biol. 6 (2005) R44.

[2] P. Gordon, Pokay, https://github.com/nodrogluap/pokay, ???? Accessed: 2023-6-7.

[3] M. Urban, A. Cuzick, J. Seager, V. Wood, K. Rutherford, S. Y. Venkatesh, J. Sahu, S. V. Iyer, L. Khamari, N. De Silva, M. C. Martinez, H. Pedro, A. D. Yates, K. E. Hammond-Kosack, PHI-base in 2022: a multi-species phenotype database for Pathogen-Host interactions, Nucleic Acids Res. 50 (2022) D837–D847.

[4] Y. Lin, Y. He, Ontology representation and analysis of vaccine formulation and administration and their effects on vaccine immune responses, J. Biomed. Semantics 3 (2012) 17.

[5] Y. He, H. Yu, E. Ong, Y. Wang, Y. Liu, A. Huffman, H.-H. Huang, J. Beverley, J. Hur, X. Yang, L. Chen, G. S. Omenn, B. Athey, B. Smith, CIDO, a community-based ontology for coronavirus disease knowledge and data integration, sharing, and analysis, Sci Data 7 (2020) 181.

*Corresponding author.

✉ m.iseminger@alumni.ubc.ca (M. Iseminger)

🆔 0000-0002-0548-891X (M. Iseminger); 0000-0001-8236-485X (M. Z. Anwar); 0000-0002-9578-0788 (R. Cameron); 0000-0002-8844-9165 (D. Dooley); 0000-0002-1107-9135 (E. Griffiths); 0000-0001-5275-8866 (A. Sehar); 0000-0002-1342-4043 (W. Hsiao)