ORGAN AND CELL-SPECIFIC BIOMARKERS OF LONG-COVID IDENTIFIED

WITH TARGETED PROTEOMICS AND MACHINE LEARNING

Maitray A. Patel, MD, PhD(c)[1], Michael J. Knauer, PhD[2]; Michael Nicholson, MD[3];

Mark Daley, PhD[1, 4]; Logan R. Van Nynatten, MD[3];

Gediminas Cepinskas, DVM, PhD[5, 6]; Douglas D. Fraser, MD, PhD[5, 7-10]

1. Epidemiology and Biostatistics, Western University, London, ON, Canada, N6A 3K7

2. Pathology and Laboratory Medicine, Western University, London, ON, Canada, N6A 3K7

3. Medicine, Western University, London, ON, Canada, N6A 3K7

4. Computer Science, Western University, London, ON, Canada, N6A 3K7

5. Lawson Health Research Institute, London, ON, Canada, N6C 2R5

6. Medical Biophysics, Western University, London, ON, Canada, N6A 3K7

7. Children's Health Research Institute, London, ON, Canada, N6C 4V3

8. Pediatrics, Western University, London, ON, Canada, N6A 3K7

9. Clinical Neurological Sciences, Western University, London, ON, Canada, N6A 3K7

10. Physiology & Pharmacology, Western University, London, ON, Canada, N6A 3K7

Key Words: Long-COVID; COVID-19; Targeted Proteomics; Machine Learning; Organ System; Cell types, Biomarker

ABSTRACT

Background: Survivors of acute COVID-19 often suffer prolonged, diffuse symptoms post-infection, referred to as "Long-COVID". A lack of Long-COVID biomarkers and pathophysiological mechanisms limits effective diagnosis, treatment and disease surveillance. We performed targeted proteomics and machine learning analyses to identify novel blood biomarkers of Long-COVID.

Methods: A case-control study comparing the expression of 2925 unique blood proteins in Long-COVID outpatients versus COVID-19 inpatients and healthy control subjects. Targeted proteomics was accomplished with proximity extension assays, and machine learning was used to identify the most important proteins for identifying Long-COVID patients. Organ system and cell type expression patterns were identified with Natural Language Processing (NLP).

Results: Machine learning analysis identified 119 relevant proteins for differentiating Long-COVID outpatients (Bonferonni corrected P<0.01). Protein combinations were narrowed down to two optimal models, with nine and five proteins each, and with both having excellent sensitivity and specificity for Long-COVID status (AUC=1.00, F1=1.00). NLP expression analysis highlighted the diffuse organ system involvement in Long-COVID, as well as the involved cell types, including leukocytes and platelets, as key components associated with Long-COVID.

Conclusions: Proteomic analysis of plasma from Long-COVID patients identified 119 highly relevant proteins and two optimal models with nine and five proteins, respectively. The identified proteins reflected widespread organ and cell type expression. Optimal protein models, as well as individual proteins, hold the potential for accurate diagnosis of Long-COVID and targeted therapeutics.