Combining **proteomics** and **metabolomics** to *identify signatures protective* of **neurological consequences** of **post-acute SARS-CoV-2 infection**

Chen-Yang Su [1], Katerina Trajanoska [2], Vincent Mooser [2], Sirui Zhou [2]

The McGill Genomic Medicine CERC Program

[1] Quantitative Life Sciences Program, McGill University, Montreal, Quebec, Canada
[2] Canada Excellence Research Chair in Genomic Medicine, Department of Human Genetics, Faculty of Medicine and Health Sciences, Victor Phillip Dahdaleh Institute of Genomic Medicine, McGill University, Montreal, Quebec, Canada.

chen-yang.su@mail.mcgill.ca | sirui.zhou@mcgill.ca
@chenyangsu | @SiruiZhou
https://www.genomic-medicine-cerc.online/

CIHR IRSC Canadian Institutes of Health Research / Instituts de recherche en santé du Canada

BQC19 BIOBANQUE QUÉBÉCOISE DE LA COVID-19

LONG COVID WEB

Introduction

Post-acute sequelae SARS-CoV-2 (PASC) is an emerging public health concern with heterogeneous manifestations. Among PASC symptoms, long-term **neurological manifestations** are particularly debilitating. Their molecular underpinning remains poorly understood. Here, we combined **proteomics** and **metabolomics** to address this issue.

Aims

Finding **meaningful biomarkers** that explain **neurological PASC** by *integrating* different omics modalities

Methods

Proteomic and **metabolomic** data were generated for **1,234 individuals** from the **Biobanque Québécoise de la COVID-19 (BQC19)**, which followed COVID-19 patients over two years. After quality control, **4,984 proteins** and **943 metabolites** were retained for analysis. Using **689 clinical entries** from the BQC19, we defined ***broad neurological PASC (BNP)*** as individuals that were *COVID-19 positive with PASC symptoms 28 days after the first symptom onset and having answered yes to any conditions related to neurological manifestations*. **Controls** were defined as *all COVID-19 positive individuals that were not cases*. **Single biomarker logistic regression models** while adjusting for age and sex to determine the association of each protein or metabolite with BNP. **L1 regularized logistic regression (LASSO) models** were trained to predict BNP using ten repeats of five-fold cross-validation to select optimal hyperparameters. **Multi-Omics Factor Analysis (MOFA)**, a versatile and statistically rigorous generalization of PCA that produces an interpretable low-dimensional representation of the original data was used to generate latent factors. We queried the **Human Metabolome Database (HMDB)** for neurological PASC-associated proteins and metabolites to determine whether these biomarkers were within the same pathway or associated with one another.

Reference: Argelaguet, R. et al. Multi-Omics Factor Analysis—a framework for unsupervised integration of multi-omics data sets. Molecular Systems Biology 14, e8124 (2018).

Results

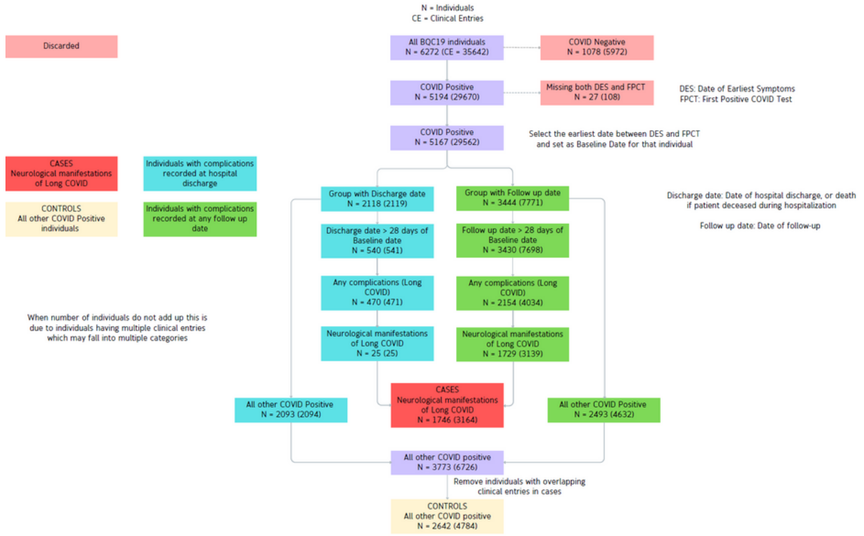

Figure 1. Flowchart showing how neurological PASC cases and controls were defined starting from Biobanque Québécoise de la COVID-19 (BQC19) data.

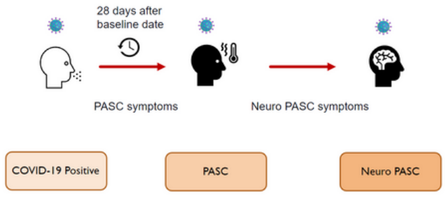

Figure 2. Cartoon diagram showing the definition of neurological PASC starting from COVID-19 positive individuals. We removed individuals that never had a positive COVID-19 test leaving only individuals that tested positive for COVID-19 at some time point. For each individual, we used the earliest date between the date of earliest symptoms (DES) or their first positive covid-19 test date (FPCTD) as the baseline date.

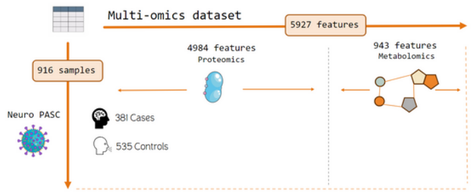

Figure 3. Schematic showing the multi-omics neurological PASC dataset which contains 381 BNP cases and 535 controls with both proteomics and metabolomics measurements.

	FDR significant biomarkers
Proteomics (N = 4,984)	1,416
Metabolomics (N = 943)	158

Table 1. Single-modality biomarker association results showing the number of false discovery rate (FDR) significant biomarkers after logistic regression (Neuro PASC ~ age + sex + single biomarker). Significance threshold was set to FDR p < 0.05.

Model Type	Features	Selected Features	Selected Lambda	Average AUC over validation folds
Baseline Model	Age, sex	2	10	59.1%
Protein Model	Age, sex, 4984 proteins	40	31.62	63.7%
Metabolite Model	Age, sex, 943 metabolites	82	17.78	62.8%
Omics Model	Age, sex, 4984 proteins, 943 metabolites	49	31.62	63.0%

Table 2. L1 regularized logistic regression results suggest that neurological PASC is difficult to predict even with proteomics and metabolomics

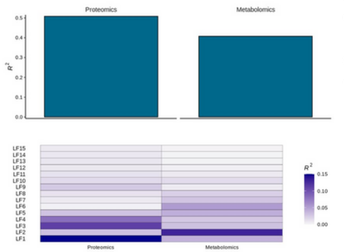

Figure 4. Multi-omics factor analysis (MOFA) identified 9 latent factors (LFs) when setting a minimum explained variance of 2% in at least one data type. The first 5 LFs were present in both omics types, indicating broad roles in multiple molecular layers. LFs 6-9 were specific to a single data modality. Cumulatively, the 9 factors explained 45% of the variation in the proteomics data and 35% in the metabolomics data.

Figure not shown. We identified phosphatidylethanolamine-binding protein 1 (PEBP1) as being protective against BNP (OR (95% CI) = 0.76 (0.66-0.87), FDR p = 0.002). Interestingly enough, previous studies suggested downregulation of PEBP1 may lead to Alzheimer's disease. Moreover, PEBP1 functions as an enzyme for 1-palmitoyl-2-docosahexaenoyl-GPE, a phosphatidylethanolamine (PE) whose impairment may lead to neurodegenerative disorders, which is concordant with our finding that increased circulating PE is associated with decreased neurological PASC risk (BNP: OR (95% CI) = 0.78 (0.68-0.90), FDR p = 0.01.

Conclusion

Here, we used an integrative bi-omics approach combining proteomics and metabolomics to provide new insight into the pathophysiological mechanisms underlying neurological PASC risk which, possibly, could point to new ways to treat this condition.

Acknowledgements

Chen-Yang Su is supported by a Canadian Institutes of Health Research Doctoral Award (FRN: 187673) and has been previously supported by an FRQS Doctoral training scholarship, a Lady Davis Institute/TD Bank Studentship Award, and a Quantitative Life Sciences PhD stipend award.