Differential Glycoprotein Profiles in Severe and Asymptomatic COVID-19

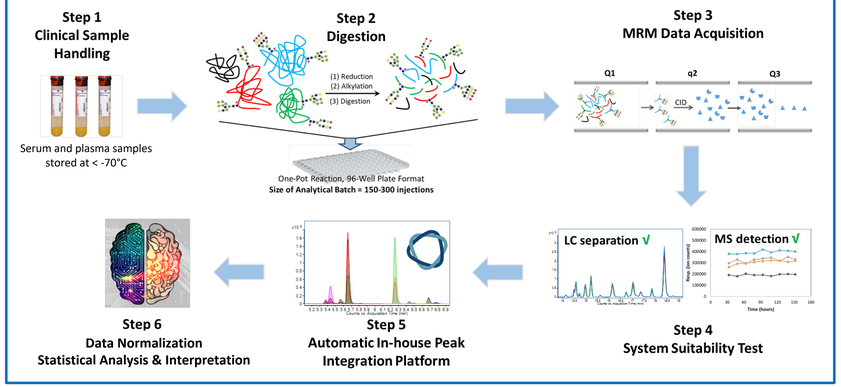
Chad Pickering[1]; Prasanna Ramachandran[1]; Gege Xu[1]; Rachel Rice[1]; Bo Zhou[1]; Hector Huang[1]; Klaus Lindpaintner[1]; Srinivasa T. Reddy[2]; Faheem Guirgis[3]; Taia T. Wang[4]; Daniel Serie[1]; Xin Cong[1]
[1]InterVenn Biosciences, South San Francisco, CA; [2]University of California, Los Angeles, Los Angeles, CA; [3]University of Florida, Gainesville, FL; [4]Stanford University, Stanford, CA

Introduction

Glycosylation is the most common post-translational modification of proteins, critically affecting their structure and function. We analyzed abundant serum and plasma proteins in 50 individuals with severe COVID-19 and in 22 subjects who had an asymptomatic course of COVID-19 using liquid chromatography and mass spectrometry (LC-MS) for high-resolution site-specific quantification of glycopeptides coupled with high-throughput artificial intelligence/neural network (AI)-powered data processing of protein glyco-isoform distributions. As additional comparisons, we included 12 individuals seropositive for common-cold coronavirus (CC), 16 patients with bacterial sepsis, and 15 healthy subjects without history of coronavirus exposure. Data on 597 peptides and glycopeptides were analyzed using supervised and unsupervised machine learning (ML) techniques.

Methods

Peptides and glycopeptides from blood serum and plasma were identified in a prior study with data-dependent acquisition. Among them, a panel of >500 glycopeptides was selected for the targeted quantitative analysis of trypsin-digested serum and plasma samples using the workflow below.

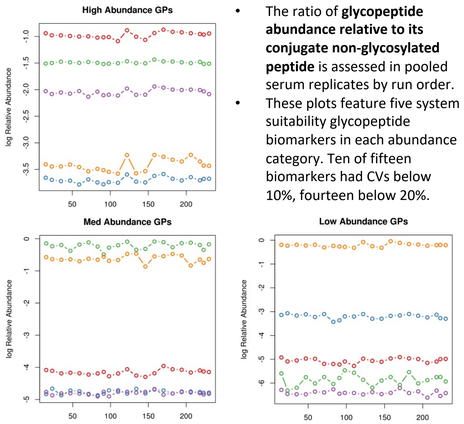


Results and Discussion

Sample Information

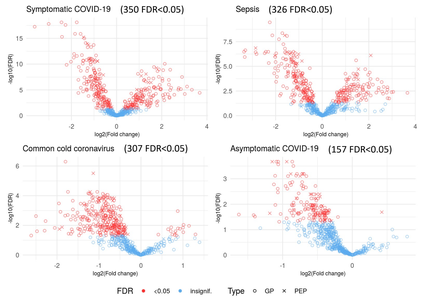
	N	Training set	Test set
Total sample size	115	86	29
Sympt. COVID-19	50	38	12
Asympt. COVID-19	22	16	6
Bacterial sepsis	16	12	4
Common cold	12	9	3
Healthy controls	15	11	4

LC/MS Quality Control Metrics

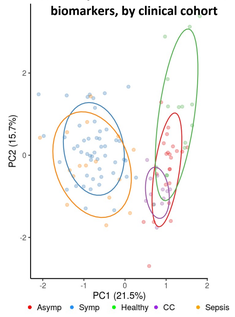


- The ratio of **glycopeptide abundance relative to its conjugate non-glycosylated peptide** is assessed in pooled serum replicates by run order.
- These plots feature five system suitability glycopeptide biomarkers in each abundance category. Ten of fifteen biomarkers had CVs below 10%, fourteen below 20%.

Glycoprotein profile comparisons between healthy subjects and 4 clinical cohorts exposed to infectious agents

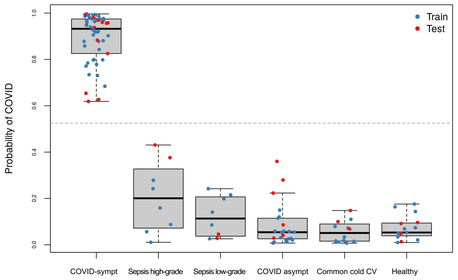


Principal Component Analysis of summary data on 597 normalized biomarkers, by clinical cohort

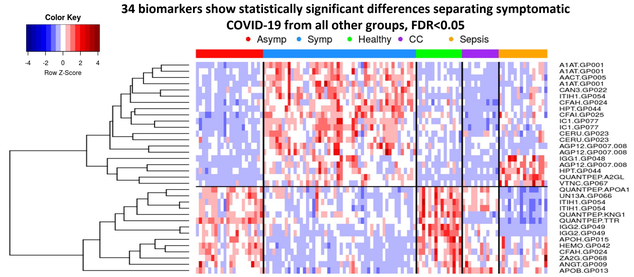


Multivariable ML Modelling

- A multivariable leave-one-out repeated cross-validated **LASSO-regularized logistic regression model** predicting probability of symptomatic COVID-19 uses 16 glycopeptides and 2 non-glycosylated peptides to achieve 100% accuracy in both training and test sets.



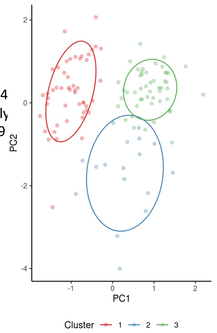
34 biomarkers show statistically significant differences separating symptomatic COVID-19 from all other groups, FDR<0.05



- **K-means clustering** was performed to predict group membership (symptomatic COVID-19, sepsis, or other), achieving classification accuracy of 96% based on a set of 34 features that statistically significantly differentiate symptomatic COVID-19 from all other cohorts at FDR<0.05 (heatmap at left).

Cluster allocations

	1	2	3
COVID+ symp	0	3	47
Other	49	0	0
Sepsis	0	14	2



- Features extracted at the glycosylation and protein levels **clearly differentiate symptomatic COVID-19 from asymptomatic/mild COVID-19** and other comparison cohorts such as bacterial sepsis and common cold (CC).
- A number of glycoproteomic biomarkers were shared between patients suffering from symptomatic COVID-19 and bacterial sepsis, likely representative of a **systemic inflammatory response**.
- A subset of 34 glycoproteomic biomarkers were **specific to severe COVID-19**, potentially providing insights into specific pathological consequences of, or a particular predisposition for, a severe course of the infection.

Conclusions

- Glycoproteomic profiles in patients with a severe course of COVID-19 infection are statistically significantly different from those who experienced an asymptomatic course. InterVenn's proprietary perspectIV™ platform, based on a high-throughput LC-MS/AI-ML approach, interrogates the blood glycoproteome at an unprecedented depth and scale, revealing a whole new domain of biology for the identification of novel biomarkers in a broad range of applications.