Proteomic Profiling of Urine From Hospitalized Patients With Severe Pneumonia due to SARS-CoV-2 vs Other Causes: A Preliminary Report

Lindsay Wilson,[1,a] Ju-Wei Chang,[1,a] Stuart Meier,[1] Tariq Ganief,[2,3] Naadir Ganief,[2,3] Suzette Oelofse,[1] Vicky Baillie,[4,5] Marta C. Nunes,[4,5] Shabir A. Madhi,[4,5,6] Jonathan Blackburn,[2,3,a] and Keertan Dheda[1,7,a]

[1]Centre for Lung Infection and Immunity, Division of Pulmonology, Department of Medicine and UCT Lung Institute & South African MRC/UCT Centre for the Study of Antimicrobial Resistance, University of Cape Town, Cape Town, South Africa, [2]Department of Integrative Biomedical Sciences, Faculty of Health Sciences, University of Cape Town, Cape Town, South Africa, [3]Institute of Infectious Disease and Molecular Medicine, University of Cape Town, Cape Town, South Africa, [4]South African Medical Research Council Vaccines and Infectious Diseases Analytics Research Unit, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa, [5]Department of Science and Technology/National Research Foundation, South African Research Chair Initiative in Vaccine Preventable Diseases, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa, [6]African Leadership in Vaccinology Expertise, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa, and [7]Faculty of Infectious and Tropical Diseases, Department of Immunology and Infection, London School of Hygiene and Tropical Medicine, London, UK

The pathogenesis of coronavirus disease 2019 (COVID-19) pneumonia remains poorly understood. The urine proteome of hospitalized patients with severe COVID-19 pneumonia, compared with severe non-COVID-19 pneumonia controls, was distinct and associated with lower abundance of several host proteins. Protein-specific machine learning analysis outlined biomarker combinations able to differentiate COVID-19 pneumonia from non-COVID-19 pneumonia controls.

Keywords. hospitalized; pathogenesis; pneumonia.

Coronavirus disease 2019 (COVID-19), caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), has claimed at least 6.87 million lives globally as of February 2023. Although vaccines, in combination with antiviral and immunomodulatory interventions, have markedly reduced the rate of hospitalization and deaths, case fatality remains significant [1]. COVID-19 is likely to become endemic, albeit with ongoing risk of severe disease in high-risk individuals, and

the risk of postacute sequelae of COVID-19 (PASC) could continue contributing to morbidity [2, 3].

Despite advances in our understanding of COVID-19, including that hyperinflammatory host immune response syndromes, including leukocyte activation syndrome, endotheliopathy, and thrombo-inflammation, plays an important role, the pathogenesis of COVID-19 remains poorly understood [4]. Urine proteomic profiling has allowed for a better understanding of host and viral [5] biomarkers associated with COVID-19 [6–11]. Such profiling is able to differentiate COVID-19 patients from COVID-19-negative individuals [7, 8], healthy controls [6, 9], those with non-COVID-19 pneumonia [6], and COVID-19 convalescing patients [9]. Classification of COVID-19 based on urine proteomic profiling is also possible according to disease severity [6, 12–14] and progression [6, 9, 13–16].

Of note, a study in the early phase of the pandemic indicated that the human urine proteome was able to differentiate patients who were hospitalized with severe COVID-19 pneumonia (ancestral variant) from non-COVID-19 pneumonia controls [6]. However, to our knowledge, no further studies in the context of other variants, and none from low- and middle-income countries, where the immunological and resource landscape is different, have been published. A preliminary study was therefore performed to compare the urine proteomic profile of hospitalized patients with COVID-19 pneumonia with the urine proteomic profile of those with non-COVID pneumonia.

Ethical approval was obtained from the University of Cape Town Faculty of Health Sciences Human Research Ethics Committee (HREC 014/2021). Hospitalized patients with acute respiratory symptoms, chest radiographic abnormalities/infiltrates, and those requiring supplementary oxygenation (therefore fitting symptoms consistent with community-acquired pneumonia) were recruited at Groote Schuur Hospital in Cape Town between April and mid-July 2021 (time frames concordant with the Beta and the Delta variant waves [17]. Patients' comorbidities were documented, and severity was graded according to the World Health Organization clinical progression scale (all participants were grade 4–6) (Supplementary Table 1) [18]. Nasopharyngeal swabs were collected, and COVID-19 status was confirmed by nucleic acid amplification tests (NAATs). Midstream urine from 20 patients, 10 with COVID-19 pneumonia and 10 non-COVID-19 pneumonia controls, was collected.

Urine was treated with 1% Triton X-100 to inactivate the SARS-CoV-2 virus, and urinary proteins precipitated and digested to peptides. The input sample was corrected and thus normalized for protein concentration between samples.

Peptides were analyzed by quantitative proteomics and mapped to human and COVID-19 protein databases. Expression analysis was performed to identify differentially abundant proteins between the groups. The proteins with the smallest *P* values were subsequently assessed for importance in random forest classification. Numerous random forest classification models were run using various combinations of the top 5 most important proteins in order to identify the best 2–3 protein biomarker combinations that could distinguish between COVID-19 cases and controls. The Benjamini & Hochberg (FDR) method was used to correct *P* values for multiple hypothesis testing. Detailed methods are provided in the Supplementary Data.

There were no significant differences between the COVID-19-positive and -negative persons when comparing gender, current smoking status, comorbidities other than HIV, and days between symptom onset and study enrollment. There was, however, a higher prevalence of people with HIV in the COVID-19-negative group ($P = .041$), and age was significantly higher among the COVID-19-positive cases ($P = .032$) (Supplementary Table 2).

Mass spectrometric analysis of the urinary proteins identified a total of 854 human proteins in the urinary proteome across all individuals. Following filtering to retain those proteins present in ≥70% of the patients in any 1 group, 286 protein groups remained. No SARS-CoV-2-specific peptides were detected in any of the patients. However, a previous study was able to detect SARS-CoV-2-specific peptides in urine [9]. The inability to detect the viral proteins in our study is most likely related to low levels of virus in the urine and minimal kidney involvement [19], although we used similar sample preparation methods and patients were recruited in the early disease stage (viral loads were variable across samples and could have contributed to the lack of detection of viral proteins in the urine).

Principal component analysis indicated relatively distinct proteins grouping in the COVID-19-positive cases, with the exception of 3 negative controls, which grouped more closely with COVID-19-positive individuals (Figure 1A). This supports the contention that COVID-19 cases and controls have distinct urine proteome signatures. Distinctive clustering was also observed in the heatmap (Figure 1B), with the exception of 1 sample. It has previously been suggested that the grouping of samples can be influenced by patient comorbidities [9], and this may have explained some of the findings. However, HIV status did not explain the clustering, which was most likely COVID-19-related, as differential expression analysis of HIV-infected persons in each group and in the whole cohort did not meaningfully change the conclusions (and the distinct clustering effect) (Supplementary Table 3).

Differential expression analysis indicated that the top 10 proteins with the lowest *P* values had reduced abundance in COVID-19-positive patients compared with negative controls (Figure 1B; Supplementary Table 3). Indeed, another study reported ~10 times more proteins with reduced abundance in COVID-19 patients compared with controls [6]. Following correction for multiple comparisons, only 3 proteins, Trefoil Factor 3 (TFF3; FDR = 0.003), Zinc Finger, Myeloproliferative and Mental Retardation–Type Containing 4 (ZMYM4; FDR = 0.042), and Fibrinogen Alpha Chain (FGA; FDR = 0.049), were significantly downregulated in COVID-19-positive cases (Figure 1C; Supplementary Table 2). Interestingly, these specific proteins were not downregulated in the COVID-19 group in a previous study [6], and we speculate that this may be a variant-specific effect.

TFF3 belongs to the Trefoil Factor Family (TFF) proteins, which are secreted by the mucosal epithelium and are thought to be involved in mucosal protection by trapping microorganisms [20]. Moreover, loss of TFF3 was linked to increased colonic inflammation and implicated in the late response to mucosal damage [21]. Downregulation of TFF3 in COVID-19-positive individuals may therefore potentially lead to a decreased ability of the host to neutralize virus and may be associated with compromised mucosal immunity. This has possible implications for mucosally delivered interventions including inhaled vaccines, especially in immunocompromised persons [22]. ZMYM4 is thought to play a role in the regulation of cell morphology and cytoskeletal organization [23], DNA damage response, and cell cycle regulation [24]. Interestingly, loss of ZMYM4 has been associated with colon cancer [25]. Downregulation of ZMYM4 in COVID-19-positive patients could therefore possibly be linked to increased DNA damage and dysregulation of the cell cycle. FGA is the alpha subunit of the coagulation factor fibrinogen, which plays an essential role in tissue healing [26, 27]. Fibrinogen is broken down to fibrin, forming the central structure of blood clots. This could explain the downregulated FGA levels found in COVID-19 patients, in whom micro- and macrovascular thrombosis is common [28]. Interestingly, loss of FGA promotes tumor growth in lung cancer [29].

Random forest analysis on the differentially expressed proteins identified the top 20 proteins associated with COVID-19 (Figure 1D). Predictive modeling was performed, and a model was selected from the best performing algorithm that achieved our desired metrics (>90% sensitivity, >90% negative/positive predictive value, and >70% specificity using 2 to 3 biomarkers). Best fit models were generated using different combinations of 3- or 2-variable combinations (Table 1). The greatest specificity, positive predictive value, and Youden score were achieved using a 3-variable combination of FGA, TFF3, and carboxypeptidase vitellogenic–like (CPVL) proteins followed by FGA, TFF3, and mesothelin (MSLN) (Table 1). Changing the probability cutoffs lowered the positive predictive values and Youden scores (Supplementary table 4). CPVL, a serine-type carboxypeptidase, was suggested to be involved in the inflammatory protease cascade and trimming of peptides for antigen presentation [30]. Interestingly, CPVL
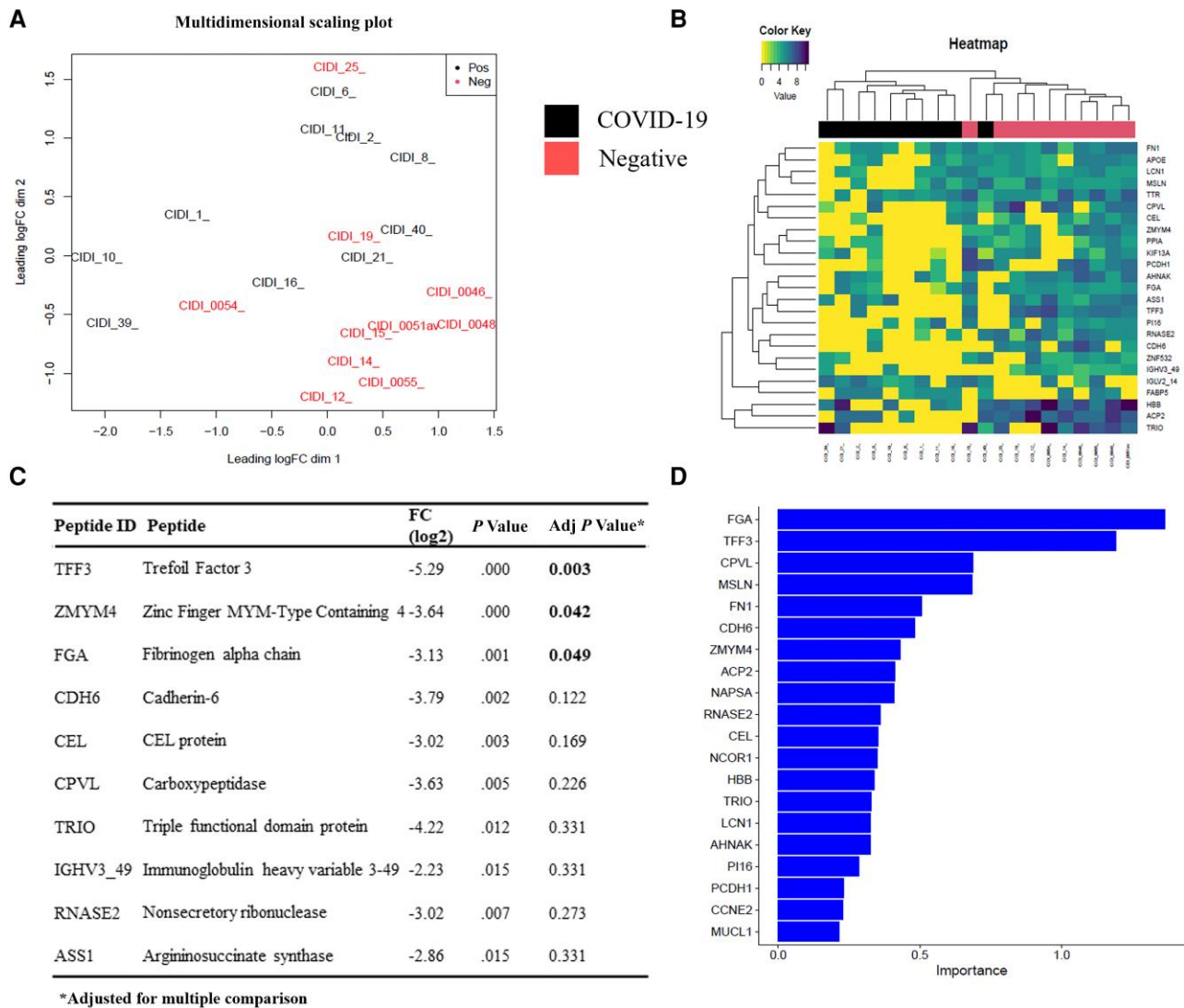
Figure 1. *A*, PCA plot showing differential clustering of samples according to protein expression levels. *B*, Heat map indicating downregulation of proteins in COVID-19-positive vs -negative patients. *C*, Top 10 differentially abundant proteins in COVID-19-positive compared with -negative patients. *D*, Random forest analysis indicating the top 20 proteins relevant to COVID-19 (in order of importance). Abbreviations: COVID-19, coronavirus disease 2019; PCA, principal component analysis.

was predicted to be the most significant risk gene for the development of severe influenza A infection [31], and lower expression of CPVL was also previously observed in severe COVID-19 compared with nonsevere COVID-19 [12]. MSLN is a tumor antigen, which is overexpressed in a variety of malignancies, including lung cancer, and is correlated with poor prognosis [32]. Interestingly, MSLN protein levels have been shown to be higher in COVID-19 patients compared with healthy controls [6].

Our results demonstrate that the 3-parameter biomarker panels identified in this preliminary study have the potential to differentiate COVID-19 from other causes of acute respiratory tract infection. However, these findings need to be confirmed in a larger study cohort. Nevertheless, we see limited utility of such an approach given the high sensitivity of antigen detection and

polymerase chain reaction–based diagnostic tools. These biomarkers do, however, shed light on the pathogenesis of COVID-19, and our work provides proof of concept that the urine proteome is an important compartment that may provide insights into better understanding the disease process and potentially prognostic biomarkers. It also confirms the limited utility of virus-specific proteins and antigen detection approaches using urine (which is different than that seen in other infections like legionella and tuberculosis where urine-based pathogen-specific antigen detection tests impact clinical practice) [33].

However, our study has several limitations. The sample size was limited (only 20 participants), but our objective was to undertake a proof-of-concept study to determine the utility of the urinary proteome in better understanding disease pathogenesis, and in the context of different and more recent variants.

Table 1. Best Fit Models of the Top 20 Most Relevant Proteins for COVID-19[a]

Metric	3 Variable Combinations				2 Variable Combinations		
	FGA, TFF3, CPVL	FGA, TFF3, FN1	FGA, TFF3, MSLN	TFF3, MSLN, CPVL	FGA, MSLN	FGA, TFF3	TFF3, MSLN
Sensitivity	0.8 (0.79, 0.81)	0.85 (0.84, 0.86)	0.88 (0.87, 0.89)	0.87 (0.86, 0.88)	0.76 (0.74, 0.77)	0.79 (0.78, 0.8)	0.81 (0.8, 0.83)
Specificity	0.95 (0.94, 0.95)	0.8 (0.78, 0.81)	0.85 (0.83, 0.86)	0.83 (0.81, 0.84)	0.89 (0.88, 0.9)	0.85 (0.83, 0.86)	0.85 (0.84, 0.86)
Positive predictive value	0.94 (0.93, 0.95)	0.81 (0.79, 0.82)	0.85 (0.84, 0.86)	0.84 (0.82, 0.85)	0.88 (0.87, 0.89)	0.84 (0.82, 0.85)	0.84 (0.83, 0.86)
Negative predictive value	0.83 (0.81, 0.84)	0.84 (0.83, 0.85)	0.87 (0.86, 0.88)	0.86 (0.85, 0.87)	0.79 (0.77, 0.8)	0.8 (0.79, 0.81)	0.82 (0.81, 0.83)
Youden	0.75 (0.73, 0.77)	0.65 (0.62, 0.67)	0.73 (0.7, 0.75)	0.7 (0.67, 0.72)	0.65 (0.63, 0.68)	0.64 (0.61, 0.66)	0.66 (0.64, 0.69)

Abbreviations: CPVL, carboxypeptidase vitellogenic–like; FGA, Fibrinogen Alpha Chain; FN1, fibronectin 1; MSLN, mesothelin; TFF3, Trefoil Factor 3.

[a]Probability cutoffs of 0.5 were used.

Moreover, our study was undertaken in a unique demographic environment and across different variants, with 21A Delta being the dominant variant at the time of sampling (May 2021 to July 2021 for this preliminary study), and therefore represents a unique and valuable data set. Second, our data are relevant to severe disease in hospitalized patients and not in ambulatory persons. However, we specifically sought to interrogate this population where better understanding and therapeutic interventions are most needed.

In summary, the urine proteome of hospitalized patients with severe COVID-19 pneumonia was distinct and associated with lower abundance of several host proteins, compared with patients with severe pneumonia due to other causes. The identified COVID-19-specific biomarkers and utility of studying the urine compartment form a foundation upon which future studies may build, thus improving our understanding of disease pathogenesis and potentially informing study approaches in future pandemics.

Supplementary Data

Supplementary materials are available at *Open Forum Infectious Diseases* online. Consisting of data provided by the authors to benefit the reader, the posted materials are not copyedited and are the sole responsibility of the authors, so questions or comments should be addressed to the corresponding author.

Acknowledgments

References

1. Bager P, Wohlfahrt J, Bhatt S, et al. Risk of hospitalisation associated with infection with SARS-CoV-2 Omicron variant versus Delta variant in Denmark: an observational cohort study. Lancet Infect Dis **2022;** 22:967–76.
2. Ssebuliba J, Nakakawa JN, Ssematimba A, Mugisha JYT. Mathematical modelling of COVID-19 transmission dynamics in a partially comorbid community. Partial Differ Equa Appl Math **2022;** 5:100212.
3. Tran VT, Porcher R, Pane I, Ravaud P. Course of post COVID-19 disease symptoms over time in the ComPaRe long COVID prospective e-cohort. Nat Commun **2022;** 13:47–56.
4. Wong LYR, Perlman S. Immune dysregulation and immunopathology induced by SARS-CoV-2 and related coronaviruses—are we our own worst enemy? Nat Rev Immunol **2022;** 22:47–56.
5. Chavan S, Mangalaparthi KK, Singh S, et al. Mass spectrometric analysis of urine from COVID-19 patients for detection of SARS-CoV-2 viral antigen and to study host response. J Proteome Res **2021;** 20:3404–13.
6. Tian W, Zhang N, Jin R, et al. Immune suppression in the early stage of COVID-19 disease. Nat Commun **2020;** 11:1–8.
7. Chavan S, Mangalaparthi KK, Singh S, et al. Mass spectrometric analysis of urine from COVID-19 patients for detection of SARS-CoV-2 viral antigen and to study host response. J Proteome Res **2021;** 20:3404–13.
8. Ye Y, Swensen AC, Wang Y, et al. A pilot study of urine proteomics in COVID-19–associated acute kidney injury. Kidney Int Rep **2021;** 6:3064–9.
9. Li Y, Wang Y, Liu H, et al. Urine proteome of COVID-19 patients. URINE **2020;** 2:1–8.
10. Li Y, Hou G, Zhou H, et al. Multi-platform omics analysis reveals molecular signature for COVID-19 pathogenesis, prognosis and drug target discovery. Signal Transduct Target Ther **2021;** 6:1–11.
11. Brogna C, Cristoni S, Petrillo M, et al. Toxin-like peptides in plasma, urine and faecal samples from COVID-19 patients. F1000Res **2021;** 10:550.
12. Bi X, Liu W, Ding X, et al. Proteomic and metabolomic profiling of urine uncovers immune responses in patients with COVID-19. Cell Rep **2022;** 38:110271.
13. Wendt R, Thijs L, Kalbitz S, et al. A urinary peptidomic profile predicts outcome in SARS-CoV-2-infected patients. EClinicalMedicine **2021;** 36:100883.
14. Staessen JA, Wendt R, Yu Y-L, et al. Predictive performance and clinical application of COV50, a urinary proteomic biomarker in early COVID-19 infection: a cohort study. Lancet Digit Health **2022;** 4:e727–37.
15. Liu Y, Song L, Zheng N, et al. A urinary proteomic landscape of COVID-19 progression identifies signaling pathways and therapeutic options. Sci China Life Sci **2022**; 65:1866–1880.

16. Wendt R, Kalbitz S, Lübbert C, et al. Urinary peptides significantly associate with COVID-19 severity: pilot proof-of-principle data and design of a multicentric diagnostic study. Proteomics **2020;** 20:2000202.

17. al Hasan SM, Saulam J, Mikami F, et al. COVID-19 outbreak trends in South Africa: a comparison of Omicron (B.1.1.529), Delta (B.1.617.2), and Beta (B.1.351) variants outbreak periods. J Infect Public Health **2022;** 15:726–33.

18. Marshall JC, Murthy S, Diaz J, et al. A minimal common outcome measure set for COVID-19 clinical research. Lancet Infect Dis **2020;** 20:e192–7.

19. Kashi AH, Karkan MF, Vaezjalali M. Urinary viral shedding of COVID-19 and its clinical associations: a systematic review and meta-analysis of observational studies. Urol J **2020;** 17:433–41.

20. Hoffmann W. Trefoil factor family (TFF) peptides. Encyclopedia **2021;** 1:974–87.

21. Hoffmann W. Trefoil factor family (TFF) peptides and their links to inflammation: a re-evaluation and new medical perspectives. Int J Mol Sci **2021;** 22:4909.

22. Heida R, Hinrichs WLJ, Frijlink HW. Inhaled vaccine delivery in the combat against respiratory viruses: a 2021 overview of recent developments and implications for COVID-19. Expert Rev Vaccines **2022;** 21:957–74.

23. Bai SW, Herrera-Abreu MT, Rohn JL, et al. Identification and characterization of a set of conserved and new regulators of cytoskeletal organization, cell morphology and migration. BMC Biol **2011;** 9:1–18.

24. Cibis H, Biyanee A, Dörner W, Mootz HD, Klempnauer KH. Characterization of the zinc finger proteins ZMYM2 and ZMYM4 as novel B-MYB binding proteins. Sci Rep **2020;** 10:1–14.

25. Moon SW, Son HJ, Chae J, et al. Expression and mutation alterations of ZMYM4 gene in gastric and colonic cancers. Appl Immunohistochem Mol Morphol **2021;** 29:570–5.

26. Moore HB, Neal MD, Ernest ME. Trauma Induced Coagulopathy. Springer; **2021.**

27. Luyendyk JP, Schoenecker JG, Flick MJ. The multifaceted role of fibrinogen in tissue injury and inflammation. Blood **2019;** 133:511–20.

28. Ryu JK, Sozmen EG, Dixit K, et al. SARS-CoV-2 spike protein induces abnormal inflammatory blood clots neutralized by fibrin immunotherapy. bioRxiv [Preprint]. **2021.** doi: 10.1101/2021.10.12.464152

29. Wang M, Zhang G, Zhang Y, et al. Fibrinogen alpha chain knockout promotes tumor growth and metastasis through integrin-akt signaling pathway in lung cancer. Mol Cancer Res **2020;** 18:943–54.

30. Mahoney JA, Ntolosi B, DaSilva RP, Gordon S, McKnight AJ. Cloning and characterization of CPVL, a novel serine carboxypeptidase, from human macrophages. Genomics **2001;** 72:243–51.

31. Li M, Chen Y, Chen T, et al. A host-based whole genome sequencing study reveals novel risk loci associated with severity of influenza A(H1N1)pdm09 infection. Emerg Microbes Infect **2021;** 10:123–31.

32. Molloy ME, Austin RJ, Lemon BD, et al. Preclinical characterization of HPN536, a trispecific, T-cell-activating protein construct for the treatment of mesothelin-expressing solid tumors. Clin Cancer Res **2021;** 27:1452–62.

33. Edelstein PH, Jørgensen CS, Wolf LA. Performance of the ImmuView and BinaxNOW assays for the detection of urine and cerebrospinal fluid *Streptococcus pneumoniae* and *Legionella pneumophila* serogroup 1 antigen in patients with Legionnaires' disease or pneumococcal pneumonia and meningitis. PLoS One **2020;** 15:e0238479.