

COVID-19

Using 'big data' to disentangle aging and COVID-19

Arthur et al. leverage different types of big data, either generated in house from cohorts of healthy aging and COVID-19, or downloaded from the ever-increasing public data archives, to disentangle the distinct cellular and proteomic mechanisms of COVID-19 and aging.

Ruth R. Montgomery and Hanno Steen

The last few decades have seen an explosive rise in the proportion of the world's population living into old age, which presents healthcare-related challenges that are epitomized by the greater susceptibility of older adults to severe COVID-19. There is also an explosion in big data and an ever-increasing richness of public data repositories archiving measurements of biological samples. Can we harness research advances in translational approaches, omic data generation, data sharing and computational techniques to advance our understanding of the aging process and its role in disease? In this issue, Arthur et al.¹ prove that we can, identifying age-independent COVID-19-specific signatures of immune subsets of CD8⁺ T cells and B cells, and elevated plasma proteins derived from liver and lung, as well as decreased skeletal muscle-derived

proteins (Fig. 1). Further, with this study, Arthur et al. illuminate a general path for more such advances across other fields.

The key strategy here is the combination and integration of traditional clinical laboratory values, immune cell subsets, plasma proteomic profiles and existing big data resources to identify relevant pathways that distinguish effects of aging from COVID-19 (moderate from severe) and other respiratory illnesses. By comparing healthy populations across different age groups with patients with COVID-19 and other non-COVID-19 respiratory illnesses, this study found, as has been reported before for this infection^{2–5}, that COVID-19 patients with severe disease were more likely to be older and male. In clinical laboratory testing, patients in the COVID-19 and respiratory illness cohorts had more neutrophils and lower red blood cell counts compared to

age-matched controls, while no statistical differences in the monocyte or platelet counts were observed. Other effects specific to COVID-19 included decreased albumin and calcium, as well as biomarkers of kidney function correlating with disease severity.

The authors used multiparameter mass cytometry (CyTOF) assays of primary peripheral blood mononuclear cells to identify immune cell populations and major subpopulations affected by aging and respiratory infection. Reduced proportions of naive cells in aging were noted, as has been well recognized across the studies of aging in the immune system^{6–8}, and which were further decreased with pulmonary infection (not exclusively to severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)). Both COVID-19 and other patient groups showed increased proportions of B cells compared to age-matched healthy

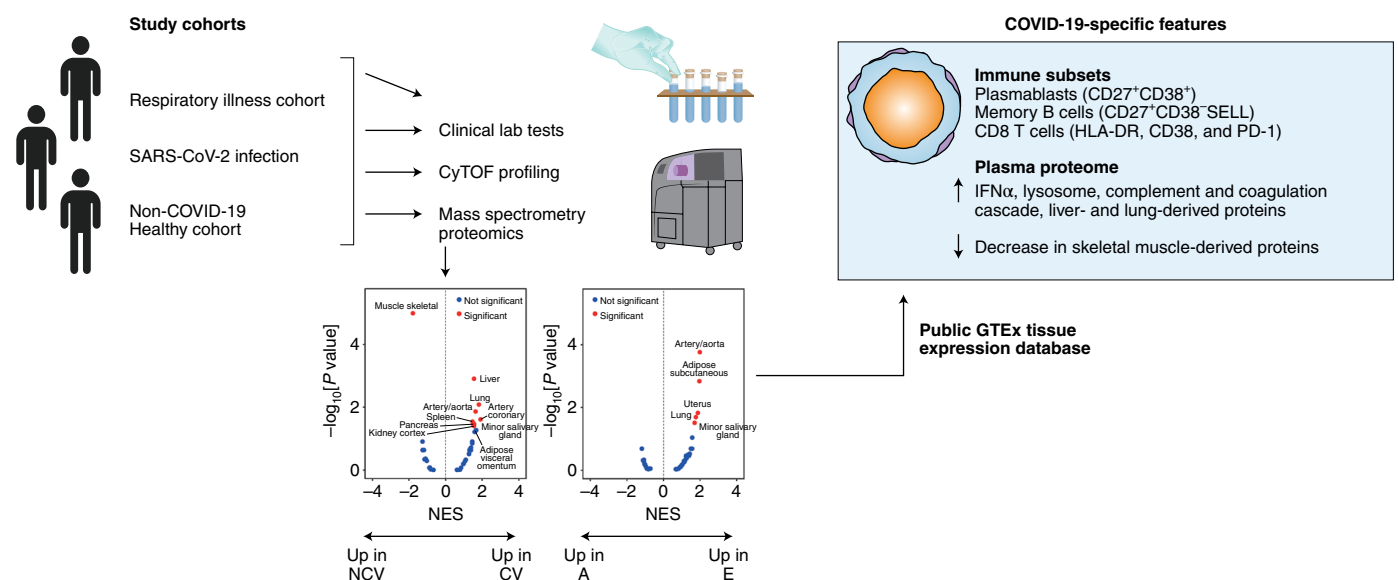


Fig. 1 | Advanced studies combine data generation and data mining in translational studies. Newly generated data from Arthur et al. was combined with public data resources to elucidate a profile of COVID-19 responses distinct from aging. NES, normalized enrichment scores; NCV, non-COVID-19; CV, COVID-19; A, 25–34 years of age; E, > 65 years of age. The graphs are reproduced from Fig. 7b in ref. ¹.

controls. Specific to COVID-19 were an increase in CD27⁺CD38⁺ plasmablasts and a decrease in CD27⁺CD38⁺SELL memory B cells. Clustering on established T-cell markers identified 12 subsets of CD4 T cells and 10 subsets of CD8 T cells with both age-specific differences and COVID-19-disease-specific changes. CD8⁺ T cells in particular showed an increase in aging of a subpopulation of CD8⁺ T cells expressing the cytotoxicity marker GZMK. Notably, compared to age-matched healthy controls, COVID-19 patients showed reduced levels of CD4 T cells, and patients with moderate COVID-19 infection showed an increase in CD8 T cells. Effector subsets of CD8 T cells expressing granzyme molecules were also increased compared to age-matched controls and other respiratory infections. COVID-19 patients specifically showed an increase in CD8⁺ T cells expressing HLA-DR, CD38 and PD-1, which the authors suggest may arise from effector memory subpopulations. CyTOF profiles differentiate 11 subsets of NK cells which differ between healthy and infected (not exclusively to SARS-CoV-2). One NK cell cluster was found to decrease with age (CD56⁺CD57⁺SELL⁺ cluster). Analysis of myeloid lineages showed decreases in classical myeloid cells with a cluster of HLA-DR^{low} myeloid cells in pulmonary infection (not exclusively to SARS-CoV-2), suggesting an immunosuppressive population. The overall implications of these differences in immune cell populations are to disentangle COVID-19-specific changes from effects of other infections and of aging. Recognizing these specialized cell subsets highlights their roles and suggests pathways critical to advance our understanding of COVID-19 disease mechanisms and responses.

To map the plasma proteomes from the COVID-19 and the healthy aging cohorts, the aptamer-based SOMAscan technology was used to detect and quantify ~4,700 proteins. Since different plasma types had been collected for the two cohorts used in this study, a direct comparison was not possible. Instead, the authors devised a strategy to first identify proteins that change due to COVID-19 or aging. In a subsequent step, the true — that is, age-adjusted — COVID-19-associated differences in the plasma proteome were mapped. After removal of proteins that showed aging-associated changes, this approach identified 337 up- and 421 downregulated proteins, respectively, that change specifically in response to COVID-19. The importance of accounting for age as a confounder was underscored by the fact that the most upregulated pathways observed in COVID-19 patients — namely, matrix and

extracellular matrix glycoproteins — were also identified as being upregulated with age.

The subsequent Gene Set Enrichment Analysis (GSEA) identified interferon alpha response, lysosome, complement and coagulation cascade, and IL2/STAT6 and IL6/STAT3 signaling as being COVID-19-specific upregulated pathways, even after accounting for age as a confounder. This age-adjusted analysis suggests pathways for investigation of biomarkers or interventions truly focused towards COVID-19.

Not all classical plasma proteins were represented in the current iteration of the SOMAscan platform, and hence it was not surprising to note that many of the plasma proteins previously described in the COVID-19 context^{9–12} were not recapitulated by Arthur et al. While missing out on certain bona fide plasma proteins, the current iteration of the SOMAscan platform is an excellent tool to identify a wide range of tissue-specific proteins¹³, and hence their aging- and COVID-19-associated dysregulation could be readily mapped in this study. To this end, Arthur et al. take advantage of the GTEx tissue expression database, a public RNA-seq resource, to identify the likely tissue sources of the proteins in the plasma identified by SOMAscan. For example, the COVID-19 patients showed a significant increase in liver- and lung-derived proteins, and a concomitant decrease in skeletal muscle-derived proteins, while aging was associated with increases in arterial and subcutaneous adipose tissue proteins. This use of big data repository information enriches the current study and emphasizes the benefits of FAIR principles (the findability, accessibility, interoperability and reuse of digital assets)¹⁴ in support of their conclusions.

While this is an elegant and informative study, no one study can address all the questions we wish to answer. It is not yet clear which of the specific immune cell and protein changes identified may directly influence COVID-19 disease severity, which mechanisms may mediate these effects and whether new therapeutic approaches may be identified. Of critical importance, the authors excluded obesity from their analysis. Although understandable, as removal of this potential confounder markedly reduced the complexity of the analysis, this will be an essential gap to address, as obesity is recognized as a critical susceptibility factor to severe COVID-19 (ref. ⁵). Future proteomic studies will have to include carefully designed age-, sex- and body mass index (BMI)-matched cohorts. In addition, plasma proteomics itself is notoriously complex due to the extreme differential of protein abundances in plasma, clouding detection of the thousands of tissue leakage proteins.

As such, SOMAscan, which is able to reliably detect >4,700 proteins, has advanced into a veritable plasma proteomic platform to detect these tissue leakage proteins in particular. Notably, the non-plasma-focused selection of detectable proteins in SOMAscan may thus underrepresent essential aspects of the immune system, the modulation of which is a key role of blood. Thus, sample-sparing liquid chromatography–mass spectrometry-based plasma proteomics and/or antibody-based cytokine assays should be considered for a comprehensive and immune-relevant plasma proteome mapping, with the additional advantage that very little additional sample is needed for these assays. Additional in-depth proteomic profiles across translational studies will also enrich data resources for comparison.

In summary, the integrated approach presented in this paper, including patient disease status and clinical laboratory data, newly generated in-depth multidimensional data and the incorporation of existing big data resources, is a real trifecta and paves the road for best practices across translational studies far beyond COVID-19 and aging studies. □

Ruth R. Montgomery¹   and Hanno Steen^{2,3}

¹Department of Internal Medicine, Yale School of Medicine, New Haven, CT, USA. ²Department of Pathology, Boston Children's Hospital and Harvard Medical School, Boston, MA, USA. ³Precision Vaccines Program, Boston Children's Hospital, Boston, MA, USA.

 e-mail: ruth.montgomery@yale.edu

Published online: 14 June 2021
<https://doi.org/10.1038/s43587-021-00078-8>

References

- Arthur, L. *Nat. Aging* <https://doi.org/10.1038/s43587-021-00067-x> (2021).
- Mathew, D. et al. *Science* **369**, eabc8511 (2020).
- Laing, A. G. et al. *Nat. Med.* **26**, 1623–1635 (2020).
- Lucas, C. et al. *Nature* **584**, 463–469 (2020).
- Palaiodimos, L. et al. *Metabolism* **108**, 154262 (2020).
- Thome, J. J. et al. *Cell* **159**, 814–828 (2014).
- Shaw, A. C., Goldstein, D. R. & Montgomery, R. R. *Nat. Rev. Immunol.* **13**, 875–887 (2013).
- Nikolich-Zugich, J. *Nat. Immunol.* **19**, 10–19 (2018).
- Messner, C. B. et al. *Cell Syst.* **11**, 11–24 (2020).
- Stukalov, A. et al. *Nature* <https://doi.org/10.1038/s41586-021-03493-4> (2021).
- Shu, T. et al. *Immunity* **53**, 1108–1122 (2020).
- Shen, B. et al. *Cell* **182**, 59–72 (2020).
- Filbin, M. R. et al. *Cell Rep. Med.* **2**, 100287 (2021).
- Wilkinson, M. D. et al. *Sci. Data* **3**, 160018 (2016).

Acknowledgements

Our laboratories are supported in part by US National Institutes of Health (NIH)/National Institute of Allergy and Infectious Diseases (NIAID) Human Immunology Project Consortium (HIPC) awards U19AI118608 and U19 AI 089992, and awards to R.R.M. (AI127865, AI142624, DA043337 and AG055362) and to H.S. (AI 152179-02 and AG 071858).

Competing interests

The authors declare no competing interests.