Cell
Leading Edge

CellPress

Previews

Advancing proteomic analysis for understanding disease biology and biomarker discovery

Carlos Cruchaga[1,2,3,4,*]
[1]Department of Psychiatry, Washington University School of Medicine, St. Louis, MO, USA
[2]NeuroGenomics and Informatics Center, Washington University, St. Louis, MO, USA
[3]Hope Center for Neurologic Diseases, Washington University, St. Louis, MO, USA
[4]Knight Alzheimer's Disease Research Center, Washington University, St. Louis, MO, USA
*Correspondence: cruchagac@wustl.edu
https://doi.org/10.1016/j.cell.2025.04.040

Advances in proteomics research have enhanced our understanding of disease biology. In a recent issue of *Cell*, Malmström et al. constructed a comprehensive proteome atlas linking proteins to specific tissues and blood cells to enable tracking of pathological changes and paving the way for broader applications in plasma proteomics across diverse diseases.

Proteins are in most cases the relevant effector molecules of biological processes and central to disease pathophysiology. Therefore, there has been great interest in recent years in studying the proteome in health and diseases. In fact, the number of published proteomic studies has increased exponentially in the last 10 years. This has been possible due to recent advances in proteomic approaches that facilitate measuring thousands of proteins in thousands of samples in an easy and quantitative manner.

Proteomic studies can be very informative for understanding the biology of diseases, identifying novel biomarkers as well as potential targets for a myriad of diseases, from cardiovascular diseases to diabetes to neurodegenerative diseases.[1–3] Most of the proteomic studies have been performed in plasma or serum, as it is a tissue easy to collect and store. However, because plasma is derived from blood, all organs of the body contribute to the measured protein levels. For this reason, it is complicated to determine where the protein was produced. When studying proteins in the context of a specific disease, this can lead to biased results or difficulties in interpreting the findings.

In order to address this problem, several studies have leveraged tissue- or organ-specific transcriptome data to perform digital deconvolution[4] and determine which cells and organs contribute to the plasma protein levels. Other studies have also used cell- and tissue-specific RNA-seq data to identify organ-specific signatures and analyze how those change with age and disease.[5] However, those studies are limited, as they are not able to differentiate at the protein level the organ or tissue of origin.

In order to address those limitations, in a recent issue of *Cell*, Malmström et al.[6] constructed an extensive proteome atlas from 18 vascularized organs and eight major cell types in blood. They integrated these data with previous RNA and protein atlases to quantitatively link proteins to 29 different human tissues and blood cells. The authors integrated all these data to generate a global label score (GLS), which represents how tissue-specific a protein is. Even though only 5% of the proteins had the highest GLS (GLS = 4), most of the proteins showed some tissue specificity, as demonstrated by additional pathway analyses (Figure 1). As could be expected, the liver, due to its central role in protein metabolism, showed the highest number of tissue-enriched proteins (n = 1,392). The organ with the second most tissue-specific proteins was the brain (n = 788), suggesting that there is more protein trafficking between the blood and brain than previously expected. These studies were done in healthy individuals, where it was not expected that the blood-brain barrier would be affected. This is an important finding, as it supports the use of plasma proteomic data to study neurological traits and diseases.[2,3,5,7,8]

The study also demonstrated that the application of this approach to plasma proteomics is well positioned to identify and track proteomic pathological changes associated with diseases. Specifically, the authors applied this technique to sepsis, pancreatitis, and myocardial injury as a proof of concept. For example, by focusing on those proteins that have a high GLS for pancreas, it was possible to separate out those individuals with pancreatitis. Applying this approach, it was possible to identify individuals at risk for the other two diseases. These results indicate that this approach can be used for more than just diagnosis, as by identifying organ-specific proteins it is also possible to identify novel biomarkers for these diseases.

The protein quantification for this study was performed using mass spectrometry, which is a good approach to perform unbiased proteomic studies on a relatively small number of samples and is therefore perfectly suited to identify organ-specific proteomic signatures. However, it is important to determine if this approach and the data from this study can be applied to other platforms such as Somalogic and Olink, which are the platforms that are being using in biobank-level proteomic studies.[9] Malmström et al. did, in fact, perform some analyses using Olink data from the UK Biobank (UKBB), and they demonstrated that individuals with liver dysfunction show higher levels of 128 liver-enriched proteins and that pulmonary surfactant

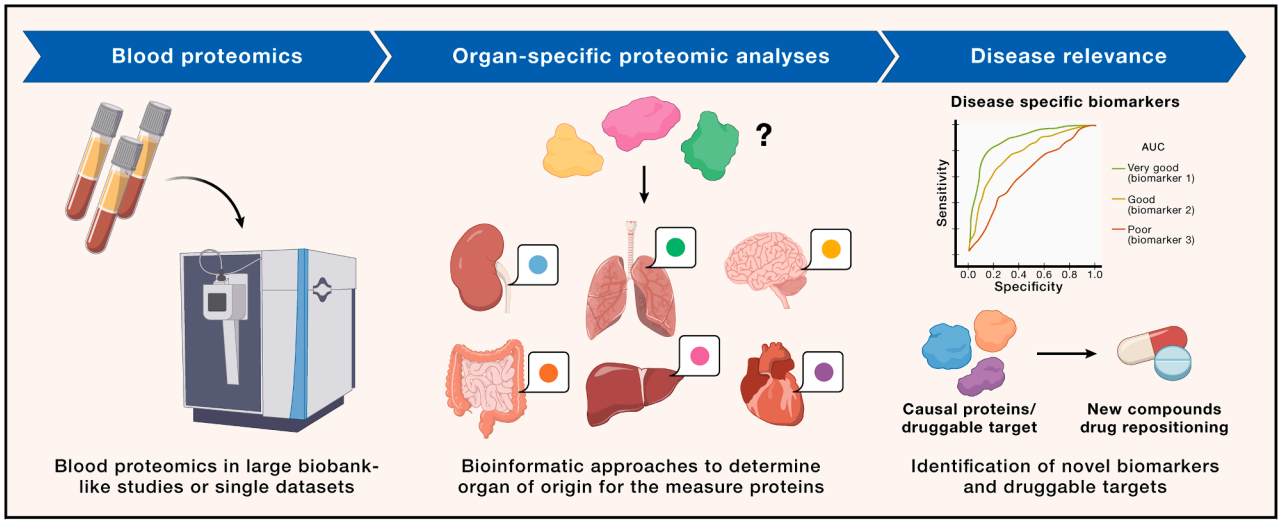

Figure 1. Plasma proteomics: A window to personalize medicine

Plasma proteomics generated from large-scale studies such as the UKBB, Finngen, or GNPC or from smaller studies can be analyzed using novel approaches, such as those presented in this study, to determine organ-specific proteins that can later be used to identify novel biomarkers as well as causal and druggable targets for complex traits and diseases.

B, a highly lung-specific protein, was significantly increased in patients with respiratory dysfunction, highly supporting that this overall approach can be applied to other platforms.

In addition, it is clear that this methodology can be applied to other diseases to deepen the understanding of plasma proteome dynamics. Newer proteomic approaches and platforms are able to measure more and more proteins (the latest Olink panel, HT1, can measure more than 5,000 proteins, Somalogic can measure 11,000 proteins, and the latest mass spectrometry methods allow researchers to measure more than 20,000 peptides) in an easy manner, democratizing proteomics to more research groups, not only those with high expertise in proteomics and mass spectrometry, as was the case 10 years ago. At the same time, there are several large-scale proteomic efforts such as the UKBB,[9] the Global Neurodegeneration Proteomic Consortium (GNPC),[2] and Finngen[10] that are generating plasma proteomics for tens of thousands of samples, as well as more individual labs generating proteomics data in their own datasets, which will translate into a plethora of new datasets that will cover many diverse diseases and traits. Therefore, there is a need to develop new analytical tools to analyze and interpret all these data. Compared to other studies, this study offers greater resolution of the tissues of origin of plasma protein levels, making it a valuable resource for disease monitoring and biomarker discovery.

DECLARATION OF INTERESTS

C.C. has received research support from GSK, EISAI, and Danaher; is a member of the scientific advisory board of Circular Genomics and owns stocks; is a member of the scientific advisory board of ADmit; and has served on the scientific advisory board for GSK and Novo Nordisk.

REFERENCES

1. Beydoun, M.A., Beydoun, H.A., Hu, Y.H., Maino Vieytes, C.A., Noren Hooten, N., Song, M., Georgescu, M.F., Fanelli-Kuczmarski, M. T., Meirelles, O., Launer, L.J., et al. (2024). Plasma proteomic biomarkers and the association between poor cardiovascular health and incident dementia: The UK Biobank study. Brain Behav. Immun. *119*, 995–1007. https://doi.org/10.1016/j.bbi.2024.05.005.
2. Cruchaga, C., Heo, G., Thomas, A., Wang, E., Oh, H., Ali, M., Timsina, J., Song, S., Liu, M., Gong, K., et al. (2025). Large-scale Plasma Proteomic Profiling Unveils Novel Diagnostic Biomarkers and Pathways for Alzheimer's Disease. Res. Sq. https://doi.org/10.21203/rs.3.rs-5167552/v1.
3. Sung, Y.J., Yang, C., Norton, J., Johnson, M., Fagan, A., Bateman, R.J., Perrin, R.J., Morris, J.C., Farlow, M.R., Chhatwal, J.P., et al. (2023). Proteomics of brain, CSF, and plasma identifies molecular signatures for distinguishing sporadic and genetic Alzheimer's disease. Sci. Transl. Med. *15*, eabq5923. https://doi.org/10.1126/scitranslmed.abq5923.
4. Teng, P.N., Schaaf, J.P., Abulez, T., Hood, B.L., Wilson, K.N., Litzi, T.J., Mitchell, D., Conrads, K. A., Hunt, A.L., Olowu, V., et al. (2024). ProteoMixture: A cell type deconvolution tool for bulk tissue proteomic data. iScience *27*, 109198. https://doi.org/10.1016/j.isci.2024.109198.
5. Oh, H.S.H., Rutledge, J., Nachun, D., Pálovics, R., Abiose, O., Moran-Losada, P., Channappa, D., Urey, D.Y., Kim, K., Sung, Y.J., et al. (2023). Organ aging signatures in the plasma proteome track health and disease. Nature *624*, 164–172. https://doi.org/10.1038/s41586-023-06802-1.
6. Malmström, E., Malmström, L., Hauri, S., Mohanty, T., Scott, A., Karlsson, C., Gueto-Tettay, C., Ahrman, E., Nozohoor, S., Tingstedt, B., et al. (2025). Human proteome distribution atlas for tissue-specific plasma proteome dynamics. Cell *188*, 2810–2822.e16. https://doi.org/10.1016/j.cell.2025.03.013.
7. Warmenhoven, N., Salvadó, G., Janelidze, S., Mattsson-Carlgren, N., Bali, D., Orduña Dolado, A., Kolb, H., Triana-Baltzer, G., Barthélemy, N.R., Schindler, S.E., et al. (2025). A comprehensive head-to-head comparison of key plasma phosphorylated tau 217 biomarker tests. Brain *148*, 416–431. https://doi.org/10.1093/brain/awae346.
8. Ibanez, L., Liu, M., Beric, A., Timsina, J., Kohlfeld, P., Bergmann, K., Lowery, J., Sykora, N., Sanchez-Montejo, B., Brock, W., et al. (2025). Benchmarking of a multi-biomarker low-volume panel for Alzheimer's

disease and related dementia research. Alzheimer's Dement. *21*, e14413. https://doi.org/10.1002/alz.14413.

9. Sun, B.B., Chiou, J., Traylor, M., Benner, C., Hsu, Y.H., Richardson, T.G., Surendran, P., Mahajan, A., Robins, C., Vasquez-Grinnell, S. G., et al. (2023). Plasma proteomic associations with genetics and health in the UK Biobank. Nature *622*, 329–338. https://doi.org/10.1038/s41586-023-06592-6.

10. Ferkingstad, E., Sulem, P., Atlason, B.A., Sveinbjornsson, G., Magnusson, M.I., Styrmisdottir, E.L., Gunnarsdottir, K., Helgason, A., Oddsson, A., Halldorsson, B.V., et al. (2021). Large-scale integration of the plasma proteome with genetics and disease. Nat. Genet. *53*, 1712–1721. https://doi.org/10.1038/s41588-021-00978-w.

Transcription factor condensates: Preventing aggregation by DNA binding

Alexandre P. Magalhaes[1] **and Denes Hnisz**[1,*]
[1]Max Planck Institute for Molecular Genetics, Berlin, Germany
*Correspondence: hnisz@molgen.mpg.de
https://doi.org/10.1016/j.cell.2025.04.034

Transcription factors can form nuclear condensates at genomic sites, and condensates are thought to enhance transcriptional activity. In this issue of *Cell*, Saad et al. suggest that DNA binding prevents rather than facilitates condensate formation of particularly aggregation-prone transcription factors.

Transcription factors (TFs) play essential roles in the conversion of genetic information from DNA into RNA transcripts. Virtually all eukaryotic TFs contain a folded DNA-binding domain that binds specific DNA sequences and one or more "effector domains" that tend to be intrinsically disordered, i.e., lack a stable secondary structure. The intrinsically disordered regions (IDRs) engage in dynamic multivalent interactions that contribute to the recruitment of various cofactors and RNA polymerase II (RNAPII).[1] Understanding how the many hundreds of TF IDRs encode specificity of interactions that ultimately converge on the same RNA polymerase machinery in the absence of a stable secondary conformation has been a major outstanding problem. A recent model suggests that TFs and their dynamic interaction partners form biomolecular condensates that concentrate them around important genomic sites (Figure 1A).[2,3] Such models have been useful for investigating sequence features hidden in TF IDRs, but the contribution of TF condensates to gene transcription and the functional roles of condensates are still debated. New insights into TF condensates are now emerging from investigating TFs that contain unusually condensation-prone IDRs.

In this issue of *Cell*, Saad et al. set out to investigate the mystery of the peculiarly long glutamine repeat in the IDR of the human FOXP2 TF.[4] A few dozen human proteins encode glutamine repeats, and the expansion of glutamine repeats is linked to the formation of pathological protein aggregates. The repeat in the IDR of wild-type FOXP2 consists of 40 glutamines, a length that causes aggregation of the HTT protein and Huntington's disease, yet curiously, there is no evidence for the aggregation of wild-type FOXP2.[4] The authors took inspiration from human genetics—namely, a known DNA-binding domain mutation, R553H, which reduces DNA binding and causes verbal dyspraxia, a rare speech disorder. Remarkably, GFP-tagged wild-type FOXP2 was found to exhibit diffuse localization in human cell nuclei, but the R553H mutant formed large nuclear condensates. Deletion of the glutamine repeat reduced condensation of the R553H FOXP2 mutant. Moreover, fusing the DNA-binding domain of the SSO7D bacterial TF to R553H FOXP2 forced DNA binding at non-native DNA sites and suppressed condensate formation.[4] As a proof-of-concept for other proteins, fusing DNA-binding domains to the HTT protein encoding pathological glutamine repeats suppressed HTT aggregation. These results appear to be in conflict with and necessitate revisiting earlier models of TF condensates.

Early TF condensate models were an attempt to reconcile observations from various imaging and genomics approaches and are supported by several lines of evidence. For example, transcription factors and their coactivators are typically visualized as mesoscale (100–500 nm) puncta in imaging experiments, and the puncta significantly co-localize with clustered enhancer DNA elements, RNAPII, and large densities of nascent RNA.[2,5] Second, mutations in TF IDRs that reduce condensation *in vitro* or in cells reduce the TF's transcriptional activity,[2,3] while mutations that enhance liquid-like features of TF IDR condensates enhance the TF's transcriptional activity.[6] In such condensate models, the TFs' DNA-binding domain contributes interaction valence through DNA binding, and the IDR contributes valence through dynamic interactions with other TFs, coactivators, or even RNA (Figure 1A). *In vitro* reconstitution experiments indeed demonstrated that mixing enhancer DNA, the OCT4 TF, and the IDR of the MED1 Mediator subunit indeed led to the formation of heterotypic condensates that require all three components.[7]