Journal of proteome research

pubs.acs.org/jpr

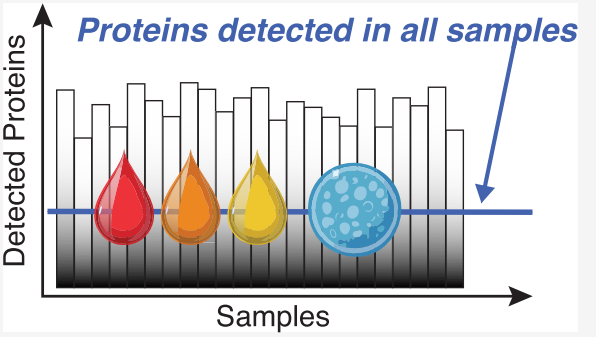
Perspective

The Circulating Proteome—Technological Developments, Current Challenges, and Future Trends

Philipp E. Geyer,* Daniel Hornburg, Maria Pernemalm, Stefanie M. Hauck, Krishnan K. Palaniappan, Vincent Albrecht, Laura F. Dagley, Robert L. Moritz, Xiaobo Yu, Fredrik Edfors, Yves Vandenbrouck, Johannes B. Mueller-Reif, Zhi Sun, Virginie Brun, Sara Ahadi, Gilbert S. Omenn, Eric W. Deutsch,* and Jochen M. Schwenk*

Cite This: https://doi.org/10.1021/acs.jproteome.4c00586

ABSTRACT: Recent improvements in proteomics technologies have fundamentally altered our capacities to characterize human biology. There is an ever-growing interest in using these novel methods for studying the circulating proteome, as blood offers an accessible window into human health. However, every methodological innovation and analytical progress calls for reassessing our existing approaches and routines to ensure that the new data will add value to the greater biomedical research community and avoid previous errors. As representatives of HUPO's Human Plasma Proteome Project (HPPP), we present our 2024 survey of the current progress in our community, including the latest build of the Human Plasma Proteome PeptideAtlas that now comprises 4608 proteins detected in 113 data sets. We then discuss the updates of established proteomics methods, emerging technologies, and investigations of proteoforms, protein networks, extracellualr vesicles, circulating antibodies and microsamples. Finally, we provide a prospective view of using the current and emerging proteomics tools in studies of circulating proteins.



KEYWORDS: *biomarker discovery, blood, plasma, serum, PTM, extracellular vesicle, microsampling, affinity, mass spectrometry, PeptideAtlas*

INTRODUCTION

The circulating proteome, consisting of the ensemble of all proteins possibly present in blood plasma or serum, holds immense potential to understand biological and pathological effects, including the promise of finding novel biomarkers that can be utilized for health and disease assessment. However, the extraordinary complexity of the circulating proteome, including various types of biomolecules and the extensive dynamic range of protein abundances spanning over 12 orders of magnitude, has posed substantial analytical challenges for decades.[1−4] As the community strives forward, we have seen remarkable advancements in affinity-based and mass spectrometry (MS)-based technologies for increasingly higher proteome coverage and throughput. This has been methodology driven by the need to address complex research questions, such as investigating the flow of information and relationships between the genome and proteome and the inherent biological variation among individuals.

As proteomic analyses of blood proteins are becoming more accessible and easier to integrate into non-proteomic studies, the community can now begin to address the remaining hurdles and opportunities. Several studies have shown that the intraindividual circulating proteome stability over time is high, but interindividual variability is significant.[5,6] These observations emphasize the importance of analyzing larger sample numbers, whether from cross-sectional studies with larger donor cohorts or through longitudinal blood collections.[7] Consequently, this underscores the necessity of solidifying capabilities for processing extensive sample sets with high precision and utilizing data from complementary analytical concepts. Methods like Mendelian randomization have identified druggable proteins with causal relationships to disease.[8] Study design and various technical and biological factors need to be considered to reduce overall variability, facilitating the discovery of physiologically relevant signatures.

Received: July 9, 2024
Revised: September 26, 2024
Accepted: September 27, 2024

Here, we aim to discuss recent technological developments in the exploration of the circulating proteome and provide a reference database of proteins currently detected through an update of the PeptideAtlas. Finally, we offer a prospective view into using the current and emerging proteomics tools in studies of proteins in plasma and serum samples to highlight current trends in the study of circulating proteins and proteoforms.

AFFINITY-BASED PLATFORMS TO PROFILE THE CIRCULATING PROTEOME

Affinity-based approaches limit surveying the proteome to a predefined set of proteins. According to Antibodypedia (https://www.antibodypedia.com), nearly five million affinity reagent products are available, covering 95% of the ∼20 000 protein-encoding genes. Over the past few years, two highly multiplexed affinity assay concepts have gained momentum for large-scale investigations.[9] The Olink proximity extension assays (PEA) exceed 5400 complementary pairs of oligonucleotide-coupled antibodies,[10] and the SomaLogic slow off-rate aptamer (SOMAmers)[11] platform now encompasses 11 000 modified aptamers. Both commercially available systems have been used across hundreds of studies, including single projects encompassing tens of thousands of samples.[12−14] While the aptamer-based assays capture proteins on beads, followed by extensive washing and hybridization of the aptamer to a DNA microarray, Olink's dual antibody assay relies on the proximal binding of two antibodies to a common target protein epitope in solution, followed by hybridization and PCR amplification of DNA barcodes with a readout by qPCR or NGS.[10,15] Similar to the latter, the recently introduced NULISA platform of Alamar Biosciences uses the concept of in-solution binding combined with capture steps of the formed immunocomplex (via the oligo and biotin tags) to stringently wash away unbound proteins to improve sensitivity.[16] All these methods require sample type and assay-specific validation efforts to achieve a selective detection of the intended target with the needed affinity to quantify low amounts. Consequently, in-house pipelines to produce affinity binders (as classical antibodies or alternative scaffolds) have been initiated to enable a more comprehensive exploration of the proteome at scale. The content of these panels has recently grown to several thousand features per sample, enabling users to uncover new biology.[17] It is worth noting that the frequency of observing all newly added proteins above the limit of detection or quantification in all samples and phenotypes will have to be determined.

A prominent topic in the community has been validating the findings from highly multiplexed affinity assays. There have been concerns about the specificity of the binders in terms of the intended proteins of interest and the availability of the affinity binders used for downstream validation analyses. Even though correlating data from one method with another is regarded as the ideal comparison, growing knowledge about other influential determinants, such as analytical performance or association with genetic variation, has made these investigations more delicate.[18,19] As shown in a recent study identifying the protein CFHR5 in venous thrombosis, one approach is to follow initial discoveries with an extensive suite of proteomics and functional experiments, including tests for causality.[20] However, targets can often only be detected on one platform or assay or require alternative concepts to infer their selectivity.[21] Testing the associations between the sequence variance of the whole genome with the portfolio of protein assay data has provided such opportunities (given that there is an association between the protein-encoding gene and its protein product).[9] In pursuit of the ground truth, complementary techniques such as mass spectrometry and biological verification should be employed, taking into account the aforementioned factors and potential preanalytical biological influences (see below).

LARGE-SCALE COMPARATIVE ANALYSES OF AFFINITY-BASED PLATFORMS

There has been a growing interest in comparing the data originating from large-scale affinity-based analyses.[18,19] In the most recent and extensive analysis, Eldjarn et al.[22] invested in the concordance between protein profiles and their genetic associations using the Olink and SomaLogic platforms. They found 576 proteins with high correlation and cis-pQTLs for the intended target (tier 1) and another 515 proteins that had a lower correlation but cis-pQTLs on either platform with a lack in protein level correlation (tier 2). As shown in Figure 1

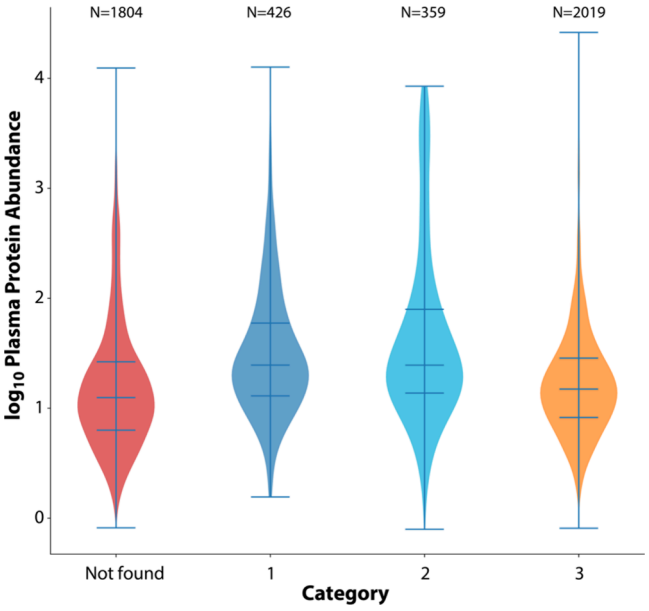


Figure 1. Protein abundance and confidence of detection by affinity assays. The boxplots show the MS-based abundance levels from the 2023 Human Plasma PeptideAtlas build and proteins classified by Eldjarn et al.[22] into confidence tiers (categories) when comparing data from Olink and SomaLogic assays. Proteins with higher support from cross-platform evidence (tier 1 and 2) have a higher abundance than proteins of lower confidence (tier 3) or those that were not detected by these affinity assays but were part of the 2023 Human Plasma PeptideAtlas build (Not found).

(74%—426/576—of proteins in tier 1 and 70%—359/515—of tier 2 were listed in the 2023 PeptideAtlas build. Compared to tier 1, proteins either lacking a cis-pQTL (tier 3, Kolmogorov−Smirnov $p = 4 \times 10^{-15}$) or those being listed in the build of 2023 were generally of lower abundance ($p = 6 \times 10^{-24}$). Many of the proteins detected at low abundance in blood are of intracellular origin or membrane bound. Other low-abundant and secreted proteins are likely related to inflammatory processes, which may require a specific phenotype to increase protein levels to a detectable range. This suggests that there are about 785 (426 + 359) proteins

https://doi.org/10.1021/acs.jproteome.4c00586
J. Proteome Res. XXXX, XXX, XXX−XXX

with supportive evidence from affinity proteomics assays that could be detectable by mass spectrometry. This set of proteins offers a valuable resource for comparing the outcome of investigations conducted on different platforms. It is expected, however, that some findings may still differ due to sensitivity, PTMs, or epitope effects.[9]

MS-BASED PLATFORMS TO PROFILE THE CIRCULATING PROTEOME

In contrast to affinity-based technologies, mass spectrometry analysis does not require a predefined set of targeted proteins. However, global MS-based analysis of the circulating proteome has been hampered by challenges in sensitivity due to the high dynamic range of protein concentrations. Consequently, few studies report over 2000 proteins detected robustly across larger data sets.[23] However, in 2023, a new mass analyzer, the Astral,[24] was introduced. Initial studies presented up to today show promising results regarding proteome coverage with over 1000 proteins in neat plasma and even more extensive proteome coverage of >4000 proteins when combined with additional enrichment methods.[25,26]

In addition to the advances in MS instrumentation, automation of sample preparation with liquid-handling platforms has been rolled out to many laboratories, scaling up throughput by parallelizing previously tedious manual work.[27−30] Automated sample preparation further allowed the implementation of ISO13485 standardization and facilitated proteomics in regulated clinical laboratories.[28] Analyzing larger sample sets calls for robustness across sample preparation workflows, LC-MS platforms, and scalable computational pipelines to enable reliable identification and quantification of proteins in all samples. This demand has driven the development of novel LC systems and the adaption of analytical flow rates (e.g., high and microflow) and shorter gradients.[31−33] Combining these with more sensitive mass spectrometers and employing new MS acquisition strategies allowed for quantitative monitoring of over a hundred samples daily.[28] Especially, data independent acquisition (DIA) has shown its ability to consistently cover high numbers of quantified proteins across samples, leading to the application of this scanning mode in numerous studies. Of note, the robustness and constant high performance of the mass spectrometers themselves have substantially improved in the past few years. Here, especially the Bruker timsTOF HT series has demonstrated its endurance in several large-scale plasma proteomics studies.[34−36] Besides higher sample throughput, such approaches reduce overall costs per sample, increase the system's stability, and achieve the desired quantitative accuracy of all detected markers. Larger cohorts not only increase statistical power and are the means against biological variation, but they also allow the application of machine learning to proteomics data.[37,38] Advances and innovations in instrumentation, sample-processing concepts, and analytical approaches are expected to emerge and be adopted in the coming years.

The unique analytical challenge of handling the wide concentration range of the plasma proteome has been a constant challenge for the field. It has, therefore, been crucial to utilize technologies capable of efficiently sampling across a broad spectrum of protein abundances. These include depleting the most abundant proteins with mainly commercial solutions of antibodies bound to a solid chromatographic phase[39,40] and investigating extracellular vesicles (see section below). Like sample preparation, these methods are automatable. Regarding the cost of sample throughput, proteome coverage can be increased by fractionation.[41] More recently, an automated workflow has been introduced to compress the large dynamic range of protein abundances by exploiting the competitive binding of proteins at the nanobiointerface of superparamagnetic functionalized nanoparticles (NPs).[42] Thus far, it has enabled thousands of circulating proteins to be surveyed within hours, promising scalability to thousands of subjects.[43,44] Unlike antibody-based depletion strategies, these engineered NPs probe the organism-agnostic physicochemical properties of entire proteomes without preselecting proteins of interest when combined with unbiased LC-MS. This has started a trend, and similar enrichment strategies are entering the field, often driven by commercial vendors. In addition to enrichment and depletion, chemical treatment of serum or plasma has been proposed to enhance the number of quantified proteins. In this approach, proteins are precipitated by perchloric acid, and analysis of the supernatant has revealed up to 1300 proteins per sample. The simplicity of the precipitation allowed early automation and scaling of the analysis to process 3000 samples.[36,45] In the initial publications, the technical reproducibility of the bead-based enrichment and the precipitation methods seemed promising, and proteins linked to different disease conditions could be identified. A wider adaptation of the methods outside of the developers' laboratories will show how useful and reliable these methods will be in the community. As is the case for all highly sensitive workflows, an improved sensitivity requires an even more careful investigation of preanalytical variability to ensure confident biological or pathological interpretation. Here, untargeted MS readouts have the advantage that they should be able to detect the contribution of preanalytical variation, such as platelet proteins, and tools can be developed to ensure the quality of the results.[46]

FUTURE OF TARGETED MS-BASED PROTEOMICS

In addition to collecting all available peptide data by unbiased MS, targeted MS-proteomics offers an approach to quantify preselected peptides consistently. Thus, its applications are tailored closer to clinical applications, which mainly rely on the validation of discovery-driven findings.[47] Targeted MS can quantify proteins with high precision, sensitivity, specificity, and throughput. The technology is often combined with spike-in of stable isotope-labeled peptide standards (SIS) to obtain absolute quantification. Serial dilution of the standard will also allow for absolute concentrations to be determined. Targeting has typically been limited in sensitivity and the quantification of less than 100 proteins. However, this has changed in recent years, and multiplex SIS panels of peptides or proteins have become increasingly useful due to their ability to measure hundreds of proteins in a single assay.[48,49] Larger multiplex panels make these concepts increasingly relevant for discovery-focused applications, similar to what current highly multiplexed affinity-based platforms offer with predefined content (panels). The applicability of targeted MS-based assays for disease classification has been demonstrated as a response to the COVID-19 pandemic.[50−52] Additionally, the development of the novel hybrid high-speed Stellar MS shows the rapid advancements of the technology tailored toward future clinical applications. The Stellar MS demonstrates high reproducibility, sensitivity, and specificity, sufficient for over 1000 plasma proteins, and has been applied to targeted assays for alcohol-related liver disease biomarkers.[53]

https://doi.org/10.1021/acs.jproteome.4c00586
J. Proteome Res. XXXX, XXX, XXX−XXX

The development of user-friendly, pretitrated mixtures of SIS peptides by various vendors has made incorporating them into various workflows easier. The ease of use has increased further thanks to software advancements with intelligent acquisition methods for ion selection and on-the-flight monitoring enabled by methods such as SureQuant,[54] Pseudo-PRM,[55] TOMAHAQ (triggered-by-offset, multiplexed, accurate-mass, high-resolution, and absolute quantification),[56] Scout-MRM,[57] and MaxQuant.Live.[58] These strategies increase the mass spectrometer's ability to concentrate data acquisition time on spiked SIS target peptides, resulting in an even more sensitive detection and a higher number of targets acquired per run.

Commercially available SIS panels often lack sufficient signature peptides per protein, causing downstream quantitative biases when only one peptide is used for the absolute concentration determination. This directly affects assay accuracy when samples must be removed due to missing data or when basing the concentration determination on different peptides.[59,60] New standards based on full-length proteins or fragments can address this issue and compensate for potential biases introduced during sample preparation.[61] With growing awareness about how genetic variation can affect protein detection, these issues must be carefully addressed to fully realize the potential of targeted proteomics in diagnostic and clinical settings.

COMMUNITY DATA

Lastly, we briefly touch upon the tools that have enabled the community to follow the progress via the generated data. The general culture of MS data sharing has been well established in the proteomics community for several years,[62] primarily driven by the efforts of the ProteomeXchange Consortium of proteomics data repositories. ProteomeXchange submissions now number approximately 600 data sets per month.[63] As shown in Figure 2, there has been an increasing number of submissions from studies using plasma samples. This data sharing has accelerated the development of open-source tools and machine-learning approaches that can model many aspects of proteomics data collection with remarkable accuracy. There are growing concerns about data privacy, patient consent, and potential personal identifiability (see General Data Protection Regulation (GDPR) in Europe or the US Health Information Portability and Accountability Act).

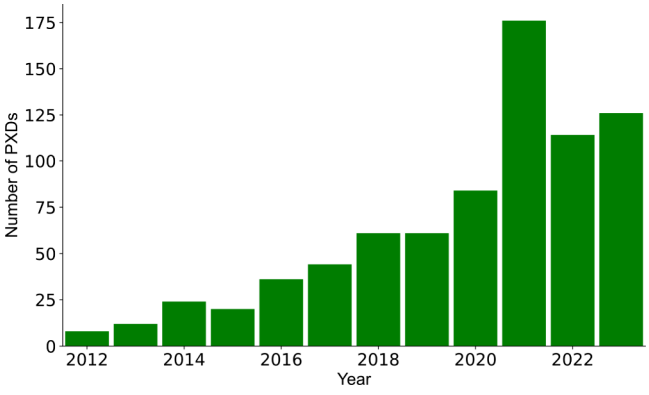


Figure 2. Number of data set submissions to ProteomeXchange over time. The submissions include only MS-based proteomics data sets. The increase in 2021 is explained by the release of the Blood Proteoform Atlas data sets.[64]

This has caused many investigators to decline the deposition of human plasma and serum data sets in public repositories. To mitigate this problem somewhat and continue to foster some degree of availability of such valuable data, there are efforts underway at PRIDE[65] and MassIVE[66] to develop controlled access data repositories, wherein data sets with public-access concerns may be deposited with gated access in the hope that the data sets can still be used with proper justification and safeguards. In addition, many high-quality data produced for or by non-academic institutions have not been shared with the community, limiting the community's possibilities to learn more about proteins in circulation.

Apart from the MS-based proteomics, community efforts around the PEA platforms have been started with the SCALLOP (http://www.scallop-consortium.com) and other consortia using this assay. These consortia are not public repositories, but new members are welcome to share data and benefit from shared data. Current data repositories are often cohort-centric, and public portals are being modified to host affinity proteomics alongside other omics data types. Thus, there will likely be many studies, like the UK Biobank,[22,67,68] where the community can obtain affinity proteomics data.

Coordinated access to affinity data would greatly benefit the research community, as demonstrated by reanalyzing MS-based proteomics data available through public repositories. A first guideline for submitting affinity proteomics data to PRIDE (PRIDE-AP) has been suggested.[69] This is particularly valuable for discovery-driven research, where there is still a limited content overlap among different platforms, which will change with technological advancements. For instance, of the 4608 proteins listed in the 2023 Human Plasma Proteome PeptideAtlas, 44% are covered by the 5416 proteins of the Olink Explore HT and 74% by the 9655 proteins of the SomaScan 11k Assay.[70,71]

2023 BUILD OF THE HUMAN PLASMA PROTEOME PEPTIDEATLAS

The Plasma PeptideAtlas has been providing a periodic snapshot[2,3,72,73] of the human plasma and serum proteomes based on public data-dependent acquisition (DDA) MS data sets since the first plasma build in 2005[74] as part of the HUPO Plasma Proteome Project (HPPP).[4] Fifteen years ago, data sets had to be solicited directly from producers since public repositories were in their infancy. Now, data sets are downloadable from public data repositories such as ProteomeXchange,[62] converted to mzML, and processed through a uniform pipeline of sequence database searching and postsearch validation with the Trans-Proteomic Pipeline.[75] Of note, there are around 125 queries per week for the Plasma PeptideAtlas build.

The 2023-04 build (serum and plasma combined) comprises 113 data sets, a 63% increase over the previous 2021-7 build, and yielded 108 million peptide-spectrum matches (PSMs), an 83% increase since the previous build. The number of distinct peptides detected increased by 27% to 0.26 million, whereas the total number of canonical proteins increased by a mere 213 to 4608. Canonical proteins require at least two non-nested uniquely mapping peptides (≥9 amino acids) that together extend at least 18 amino acids, as described in the HUPO Human Proteome Project (HPP) mass spectrometry data interpretation guidelines.[76] These 4608 proteins mapped to the

https://doi.org/10.1021/acs.jproteome.4c00586
J. Proteome Res. XXXX, XXX, XXX−XXX

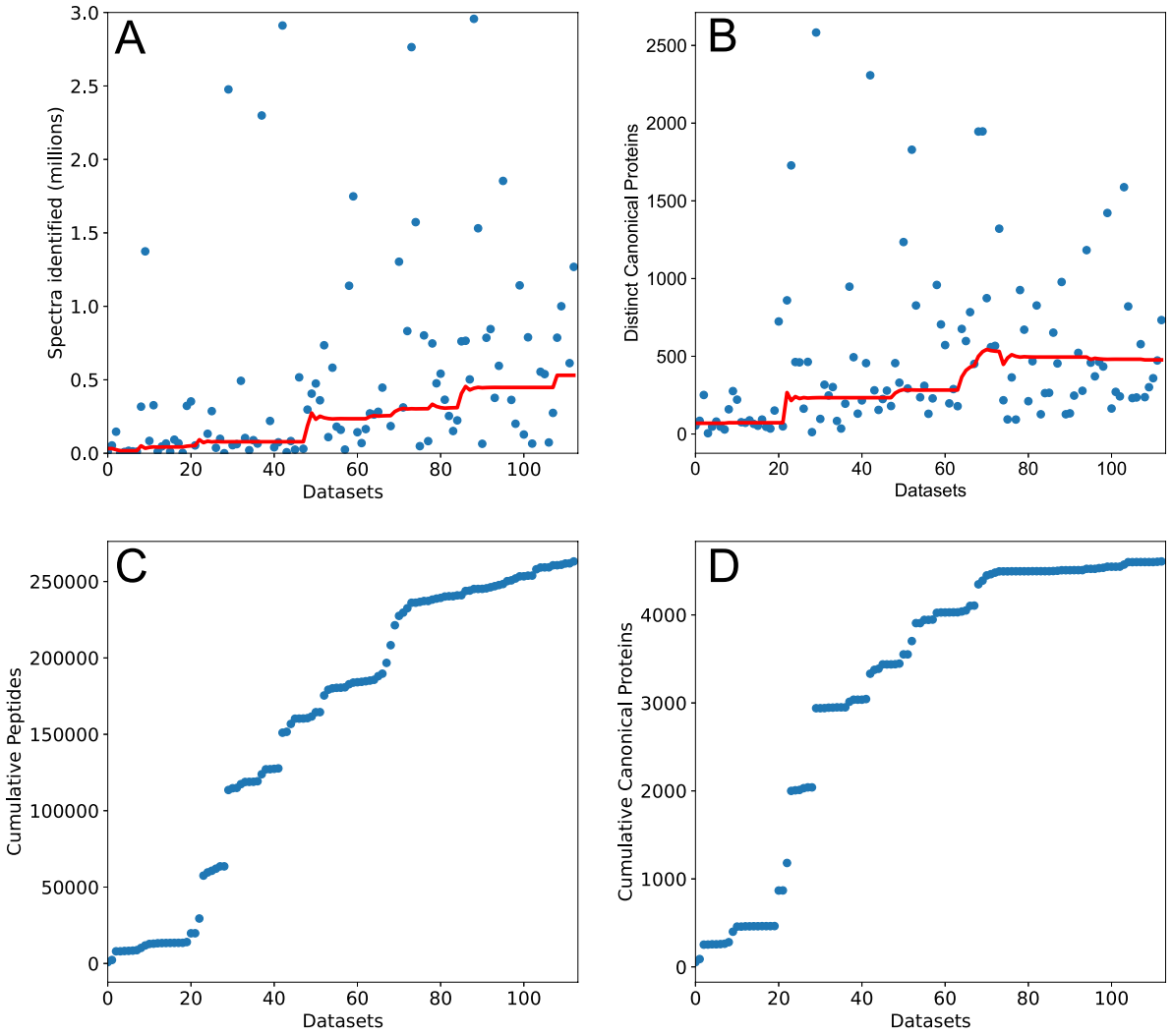


Figure 3. Trends from the 2023-04 build of the Human Plasma PeptideAtlas. (A) Number of identified spectra over time as more data sets are added. The axis is truncated to omit data sets with very high spectrum counts. A median-filter trend indicates an increase in the typical number of identified spectra per data set from 50 000 15 years ago to half a million spectra today. (B) The number of distinct canonical proteins identified over time as more data sets are added to the PeptideAtlas build. The typical number of canonical proteins detected has risen from 100 to 200 in early data sets to 500 proteins today in DDA. (C) The number of cumulative peptides in the Plasma PeptideAtlas continues to increase. The top axis provides the approximate years when data sets were added to the Plasma PeptideAtlas. (D) The number of proteins added per data set has slowed substantially, increasing by only 200 proteins despite the recent addition of nearly 50 million PSMs.

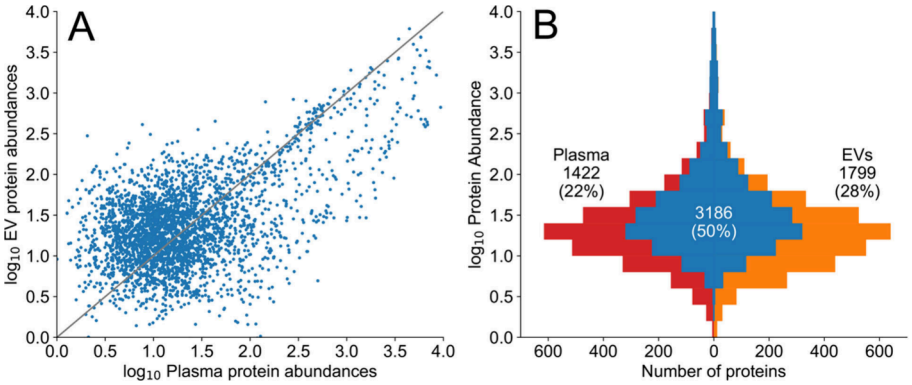


Figure 4. Comparison of protein abundances in plasma/serum samples versus EV samples. (A) Protein abundance estimated from $\log_{10}$ PSM counts for proteins seen in both PeptideAtlas builds. Most proteins correlate between plasma/serum quite well. Still, a noticeable population of proteins has a much higher abundance in plasma/serum than in EVs, while the opposite is not observed. (B) Overlap in protein identifications (blue) between plasma/serum build (red) versus EV build (orange), displayed as a function of estimated abundance ($\log_{10}$ PSM counts). 50% of proteins are seen in both builds, with ~25% unique to each build.

https://doi.org/10.1021/acs.jproteome.4c00586
J. Proteome Res. XXXX, XXX, XXX−XXX

core set of 19 778 PE1-4 core human protein entries in the neXtProt[77] resource represent 23% of the human proteome, with various isoforms and immunoglobulin variation not counted in these estimates. At the stringency of the PeptideAtlas build, only a single decoy protein achieved canonical status, yielding an estimated false discovery rate of <0.04%. All proteins, peptides, and individual spectra supporting the identifications may be browsed at the PeptideAtlas website beginning at the top page for this build at https://peptideatlas.org/builds/human/plasma/.

Figure 3 provides a view of trends over the lifetime of the Plasma PeptideAtlas including trends in the number of spectra identified per data set and the number of distinct canonical proteins per data set. Although modern data sets generate substantially more spectra that can be identified, the overall trend in the number of identified proteins is modest. The range is still quite large, driven primarily by the amount of fractionation performed in a study. Here, technological developments are needed to enable in-depth high-throughput analysis.

The rapid technological advancements in terms of higher proteome coverage in MS-based proteomics, especially facilitated by the combination of enrichment and precipitation methods and the novel Astral mass spectrometer, will continue to enrich the number of identifications. This will likely also be reflected in future builds of the PeptideAtlas.

In 2021, PeptideAtlas introduced a build comprising only data from extracellular vesicles extracted from blood.[2] The 2023-04 build of the Human Extracellular Vesicle PeptideAtlas expands on this to include 33 PXDs yielding 10 million PSMs, 181 045 distinct peptides covering 4985 canonical proteins, just 377 more canonical proteins than the plasma build. Relative abundances of plasma and EV proteins are crudely estimated based on the number of PSMs that map to each protein and are compared to each other in Figure 4A. With a similar number of proteins overall in the two builds (although a factor of 10 more PSMs in the plasma build), general abundances correlate well for many proteins. Figure 4B shows the overlap in protein identifications in the two builds as a function of abundance, illustrating the overabundance of classical plasma proteins in EV data. A very high overall overlap is evident in the upper 2 orders of magnitude, but only ∼50% overlap is observed in the lower 2 orders of magnitude.

POST-TRANSLATIONAL MODIFICATIONS—BEYOND THE CANONICAL PROTEOME

In addition to canonical proteins, the landscape of post-translational modifications (PTMs) offers a biologically relevant layer to characterize the circulating proteome. PTMs affect proteins in many ways, including solubility, structure, ligand binding, localization, stability, secretion, and activity. With more than 400 known modifications, LC-MS is a powerful technology to study a plethora of PTMs, such as glycosylation, phosphorylation, methylation/acylation, and ubiquitination.[78] An additional layer of complexity arises from the crosstalk between PTMs, leading to intricate combinations of modifications fine-tuning biological outcomes. Historically, most investigations only focused on one type of PTM, but more recently, PTM combinations and crosstalk have been characterized and studied.[79,80] Here, we highlight two types of PTMs that are found in the plasma proteome: glycosylation and phosphorylation.

While it is well-known that changes in glycosylation are associated with normal and disease-related processes, analytically characterizing those changes, including both the glycan/glycoprotein identity and their abundance, can be challenging.[81] Recent advances to address some of these challenges include (i) the development of chemical or biochemical tools to enrich glycoproteins or glycopeptides from complex mixtures, (ii) methods to resolve the site and families of glycosylations, (iii) approaches for targeted glycoproteomic analysis, and (iv) advancements in bioinformatics to improve the confidence in glycan/glycopeptide assignments.[82−85] The increased interest in glycoproteomics has resulted in some impactful observations, particularly for biomarker discovery as outlined below.

Many cancer biomarkers used in the clinic are glycoproteins, including carcinoembryonic antigen (CEA), CA125 (the repeating peptide epitope of the mucin MUC16), prostate-specific antigen (PSA), and alpha-fetoprotein (AFP).[86−88] As MS-based methods for glycan characterization evolved, extensive characterization of these glycoproteins revealed highly heterogeneous structures, sometimes with 10s to 100s of different glycoforms present, such as with CEA and PSA.[89,90] In a recent review, Chernykh et al. highlighted how advancements in LC-MS have enabled a new era of structurally resolved glycoproteomics, where researchers can now routinely assign thousands of N-linked and/or O-linked glycan structures.[91] Such methods have made it possible to explore the correlation in expression levels between parent proteins and their glycoforms, leading to the correlation of specific glycoforms and phenotypes. In a proof-of-concept study investigating genetically defined glycosylation diseases, subtle (but known) changes in site-specific N-linked glycans have been used to stratify patients according to their specific genetic defects,[92] highlighting the biological significance of glycoproteomics. Multiple-reaction monitoring (MRM) was used to assess the abundance of 159 predefined glycopeptides in healthy human plasma, while it has also been utilized to map intra- and interprotein glycan correlations. While in this study such correlations were used to stratify age− and gender−glycan relationships, the same principles could be applied to biomarkers for disease detection. Interestingly, the ability to characterize changes in specific glycoform levels in circulation likely results from changes in glycan-processing enzymes' expression, which are often dysregulated during disease progression.[88]

Another emerging opportunity for applying glycoproteomics is improving predictions for functionally relevant and disease-subtyping efforts. Recent studies in ovarian, prostate, kidney, pancreatic, and liver cancers illustrate the contribution of glycoproteins to the prediction of clinical outcomes, the severity of disease progression or risk, identifying specific pathological subtypes, or monitoring treatment pathways.[93−97] Although most current efforts utilize tissue as their primary source material, we wanted to highlight opportunities to build off this foundation and extend the results into the circulating proteome. Of note, early-stage diagnostic companies are utilizing these glycosylation changes to build blood-based diagnostic tests for early disease detection.

Like protein glycosylation, phosphorylation has also been observed in the blood proteome. While there are many mechanisms under which these modified proteins appear in circulation, three mechanisms include Golgi-mediated phosphorylation to actively secreted proteins, secreted kinases that

https://doi.org/10.1021/acs.jproteome.4c00586
J. Proteome Res. XXXX, XXX, XXX−XXX

modify proteins extracellularly, and leakage proteins from cell death or related events.[98,99] Several large-scale discovery studies have helped catalog some of these extracellular and circulating phosphoproteins, revealing interesting mechanisms related to disease progression, treatment, and/or monitoring.[100,101]

A major proteogenomics project is the National Cancer Institute-sponsored Clinical Proteomics Tumor Analysis Consortium (CPTAC), now in its 17th year. Multiomics analyses of drivers and targets for translation to oncology have been published for a broad array of cancers. For example, clear cell Renal Cell Adenocarcinomas account for 75% of kidney cancers. Clark et al.[102] identified remarkable features of 3p loss, translocations, and biallelic mutations of VHL (nearly universal) in combination with mutations of one of four tumor suppressors coded on chromosome 3p (PBRM1, BAP1, SETD2, KDM5C). Phosphoproteomic analyses showed transcription factors specifically involved in PI3K-mTOR or HIF-2alpha regulatory pathways, cell cycle, or angiogenesis. Many findings with proteomics or phosphoproteomics could not be predicted from DNA or RNA analyses. Kinase targets for many chemotherapy agents approved by the FDA or in clinical trials were identified. Four subtypes of ccRCC were identified that could guide use of immunotherapy regimens: CD8+ inflamed, CD8− inflamed, metabolic immune desert, and VEGF immune desert. Further work reported features associated with aggressiveness/poor survival of ccRCCs, including UCHL1 as a biomarker of survival, and characterized rare types of kidney tumors, including papillary, chromophobe, angiomyosarcoma, and oncocytomas.[96]

Similar studies have been published on lung adenocarcinomas (LUAD), pancreatic ductal adenocarcinomas (PDACs), and many other cancers.[103−106] An ancillary study identified circulating serum biomarkers presumably released from PDACs into the circulation.[103]

NETWORKS AND CONNECTIONS OF CIRCULATING PROTEINS

One approach to defining subproteomes is their coregulation in groups or networks, which might pinpoint distinct regulatory networks. The reason for coregulation can be found in various mechanisms, from organ or tissue leakage to the presence of circulating cells, stable protein complexes, aggregation to lipoprotein particles, EVs, or the response to specific stimuli.

Typically, tissue leakage proteins are clinical markers of organ damage (see troponin) or are used to detect cancers due to the aberrant secretion of cancer-associated proteins, such as CA125, for early detection of ovarian cancer. Next to the disease context, tissue leakage proteins can originate from biological processes. However, due to their relatively low abundance, tissue specificity, and less consistent presence, detecting a wider variety of tissue leakage proteins has been very limited to severe phenotypes such as COVID-19.[107] Using isoelectric focusing of peptides and MS-based readout, a recent study reported placenta-specific proteins released in circulation during pregnancy.[108] Combining this with genetic information (denoted *proteogenomics*) has allowed the study to assign those detected peptides with specific single amino acid variants to either the fetus or the mother. This demonstrated a cross-placenta exchange between the child and the maternal circulation system. In a study of alcohol-related liver disease, paired liver−plasma samples could be used to discern tissue damage markers and proteins involved in active signaling processes.[109]

In addition to applying differential abundance statistics to identify coregulated proteins, network-based approaches are a valuable tool to gain a deeper biological understanding of proteins or reduce collinearity for downstream statistical analysis. This is especially true for proteins with low-fold changes or high interindividual variation, which is the case for a large proportion of the proteome.[6,110,111] Coexpression network approaches, such as weighted (gene) coexpression network analysis (WGCNA), use pairwise association measures to describe the correlation relationships between analytes.[111−113] The information content of pairwise correlation analysis can be very extensive. For example, a study analyzing 1000 proteins would generate a matrix of 499 500 correlations. When combining the correlation matrix with a clustering algorithm, "global correlation maps" can be constructed, and organizational features pointing to physical connections or functional associations can be captured.[114] While correlation does not prove causation, associations with clinical parameters or anthropometric information can generate hypotheses for underlying biological mechanisms. Combining these with enrichment analysis using GO or KEGG terms can reveal further biological clues about the origin of the observed correlation cluster. Setting the limit to the included proteins, such as for the targeted affinity-based assays, or the currently technically possible markers, such as for MS-based discoveries, could reveal different findings than including the whole proteome. Another key to facilitating a more informed outcome is the sample size, with more extensive studies increasing the power to detect small effect sizes, complex signatures (i.e., many proteins comprising a unique pattern), or changes associated with cohort subsets. For example, studying serum samples of 5000 Icelanders identified regulatory networks of circulating proteins that pointed at coordinated regulation modules across human tissues associated with health outcomes and genetic variants linked to complex diseases.[115] A study analyzing 2622 samples demonstrated how to connect protein coexpression modules with clinical variables to interpret the underlying mechanisms of coexpressed proteins.[116] Another study drafted an extensive list of associations between proteins and clinical parameters in COVID-19, providing a treasure trove to interpret protein coexpression networks.[117] This can be further grown to interomic cross-sectional correlation networks, including proteomics, clinical parameters, microbiome, genetic traits, and metabolomics,[118,119] with some statistical tools such as multiomics factor analysis enabling differentiation of biological and technical factors.[120]

Technological advancements resulting in fewer missing values, higher throughput for larger studies, higher reproducibility, and tailored multiomics integration strategies will make network-centric analysis more robust, powered, and insightful. Increasing data amount and quality via higher proteome coverage will increase our knowledge about organ-centric and systemwide regulation of the circulating proteins.

THE EXTRACELLULAR VESICLE SUBPROTEOME

Extracellular vesicles (EVs) comprise a heterogeneous group of lipid bilayer particles secreted by cells and play an important role in cell-to-cell communication and cargo transportation processes. There are three main subgroups of EVs, which are classified primarily according to their sizes and biogenesis:

https://doi.org/10.1021/acs.jproteome.4c00586
J. Proteome Res. XXXX, XXX, XXX−XXX

exosomes, microvesicles, and apoptotic bodies. Exosomes are small EVs with diameters ranging from 30 to 150 nm, which are generated inside the cell in multivesicular bodies (MVBs) and subsequently released through the fusion of the MVB with the plasma membrane. Microvesicles are larger, 100−800 nm in diameter particles shed directly from the cell's plasma membrane, similar to apoptotic bodies (200 nm−5 μm) created specifically during apoptosis. EVs carry nucleic acids, such as microRNA and mRNA, lipids, and proteins, which can be transferred to the recipient cell. Studies have shown that the content of EVs differs depending on their cellular lineage and that they reflect the cells from which they originate. In addition, they offer a more accessible, cell-like dynamic range of protein concentrations. These factors make EVs highly interesting as a potential source of biomarkers and vehicles in drug delivery. From a circulating proteome perspective, EVs provide a shielded, partially hydrophobic, and nonoxidative entry mechanism into plasma. They may contain proteins (e.g., strictly intracellular proteins) not commonly detectable in circulation. They are large enough to be retained in circulation during kidney filtration. They are an integrated part of the circulating plasma proteome, albeit at relatively low concentrations ($\sim 10^{10}$ particles/mL). Studying the population of EVs in plasma (pEVs) is associated with some distinct challenges, mainly due to the heterogeneity of the particles, their relatively low abundance in plasma, and their physicochemical similarity to highly abundant, non-EV complexes in plasma, such as LDL and chylomicrons, leading to coenrichment during EV isolation. This has hampered the analysis of pEVs. Moreover, the number of identified proteins from MS analysis of pEVs is generally substantially lower than from EVs isolated from cell culture. Several enrichment techniques have been developed to enrich EVs from plasma, the most common strategies being based on ultracentrifugation, size exclusion, and immunocapture of combinations thereof. These techniques have individual advantages, but all suffer from unique challenges, including specificity, scalability, and reproducibility. Recently introduced strong anion exchange (SAX) magnetic beads have been applied for binding to phospholipid bilayer-bound particles.[121] The direct combination of protein aggregation capture (PAC) sample preparation with the magnetic particles has resulted in an automated workflow seamlessly integrating enrichment and proteomics sample preparation. Several studies have reported markers for the different EV subgroups, including tetraspanins and Alix for exosomes, actinin-4, and HSP90B1 for microvesicles. However, individual EVs are very heterogeneous, providing a challenge for selecting immunocapture targets and distinguishing between a true EV protein and a coenriched protein. Notably, the International Society of Extracellular Vesicles (ISEV) has published the "Minimal Information for Studies of Extracellular Vesicles" (MISEV https://www.isev.org/misev) guidelines since 2014. This effort should guide the standardization of protocols and reporting in the extracellular vesicle field to improve the reproducibility and quality of the EV studies. This is crucial because minor changes in protocols, such as choosing a consistent sample starting material (plasma preparations or serum), can have a significant effect on the population of EVs.

In recent years, a substantial amount of research has been put into defining subpopulations of EVs based on the presence of different surface markers, often using combinations of immunocapture and global MS analysis.[122−125] These studies provide an in-depth characterization of the diverse EVs found in plasma, identifying 1000−2000 proteins, but often at the cost of low throughput and high starting volume. These studies clearly show the potential of detecting EVs from defined cell populations, such as platelets[123,126] or specific tissue-derived EVs.[124]

(AUTO)ANTIBODY ISOTYPING

Antibodies are a key constituent of the circulating proteome, an essential part of the immune system, key players in defending against pathogens, and invaluable tools in diagnostic and therapeutic applications. Therefore, comprehensive profiling of the antibody repertoires provides crucial insights into immune responses and underlying disease mechanisms. Isotyping of antibodies can be challenging due to multiple isotypes, cross-reactivity among isotypes, and the low abundance of antibodies in some instances, making their detection and characterization arduous. Additionally, the dynamic nature of autoantibody profiles and the potential for isotype switching further complicate antibody isotyping results' interpretation and clinical relevance. Mass spectrometry is emerging to provide added information on sequence diversity and dynamics of complex antibody mixtures—information that cannot be obtained through classical immunoassays or traditional sequencing of B-cell receptor repertoires at the nucleotide level.[127]

A recent study used personalized IgA1 Fab clonal profiles from human serum and milk samples.[128] In line with previous findings,[129−131] each donor exhibits a unique repertoire of clones, and each clonal profile was considerably stable across longitudinally collected samples. The authors found that serum and milk IgA1 are dominated by a few clones, with a significant overlap arising from dimeric J-chain coupled IgA1. Conversely, the clonal repertoire of monomeric serum IgA1 was less shared, indicating the existence of two distinct sources (B cells) for the two IgA1 repertoires.[128]

MS-based antibody profiling has also shed light on the polyclonal response of individuals producing SARS-CoV-2 S-protein-targeting IgG1 clones,[132] greatly supporting serological assays,[133−136] which lack the resolution to discern the unique antibody clones. Additionally, mapping post-translational modifications (PTMs) dramatically improves our understanding of the antibody's efficacy and stability, as shown recently by linking IgG fucosylation to the severity of COVID-19[137] and secondary Dengue infection.[138]

Understanding the specific roles of antibodies will remain of high interest. For example, binding to endogenous proteins is crucial in diagnosing and treating autoimmune diseases. In the latest systematic efforts, text-mining, statistical analysis, and manual curation were used to develop a web interface that currently presents >8000 human autoantigens (AAgs) as part of the AAgAtlas portal (http://biokb.ncpsb.org.cn/aagatlas_portal/index.php[139]) containing 8045 nonredundant AAgs and 47 post-translationally modified AAgs against 1073 human diseases. Predominantly using immunoassay data, immunogenic properties of human AAgs were classified based on their genetic, biophysical, cytological, expression profile, and evolutionary characteristics. These data outline some of the hallmarks of human autoimmunity, hence presenting the value of using blood samples for the deeper characterization of the antibody repertoires and clonalities across human diseases.

https://doi.org/10.1021/acs.jproteome.4c00586
J. Proteome Res. XXXX, XXX, XXX−XXX

Table 1. Platforms for Multiplexed Analysis of Circulating Proteins (Excluding Antibodies)[a]

provider	platform specific characteristics	features	batch size	read-out system	source
Alamar Bioscience	antibody-based proximity assays	250+	80	NGS	https://www.alamarbio.com/[16]
Biognosys	depletion for ultradeep analysis	2700	NA	MS	https://www.biognosys.com/[150]
BioTechne	semiautomated microfluidic system (ELLA)	1−8	16−72	fluorescence	https://www.bio-techne.com/
Encodia	amino acid degradation	**	**	NGS	https://www.encodia.com/
Luminex***	bead-based immunoassays	1−384	96−384	flow cytometer	https://www.luminexcorp.com[151,152]
MesoScale	sandwich immunoassays	1−10	96−384	electrochemiluminescence	https://www.mesoscale.com/
Nautilus Bioscience	cycling of antibody binding	20000*	**	protein array	https://www.nautilus.bio/[153]
Olink	antibody-based proximity assays	24−5300	44−88	qPCR, NGS	https://www.olink.com[10]
Quanterix	bead or disc-based nanowell assays	1−10	88	chemiluminescence	https://www.quanterix.com/[154]
QuantumSi	amino acid degradation	**	**	own system	https://www.quantum-si.com/[155]
Resyn Biosciences	ultrafast and shallow	4000	96	MS	https://www.resynbio.com/[121]
Seer	nanoparticle enrichment	6000	40	MS	https://www.seer.bio/[156]
Siscapa	immuno-affinity capture of target peptides	200	96−384	MS	https://siscapa.com[158]
SomaLogic	aptamers with slow off-rates	11000	80	DNA array, NGS*	https://www.somalogic.com[11,157]

[a]Stars indicate whether information was recently announced (*), is currently unknown (**), or if several providers of other kits and assays are present or have been developed by users (***).

MULTIPLEXED SEROLOGY

Compared to autoantibodies, antibodies are produced in response to pathogen infections. Proteomics technologies have been developed to survey the serological profiles of infections. To exemplify their utility, a SARS-CoV-2 proteome microarray containing 966 tiled 15-mer peptides was developed to profile the B-cell epitope landscape of SARS-CoV-2 IgM and IgG antibodies in early COVID-19 infection.[140] Others constructed a SARS-CoV-2 protein array containing 18 purified viral proteins to detect the IgG and IgM antibody responses.[141] In addition, antibody profiles of COVID-19 patients and prepandemic controls were analyzed using immunoprecipitation and sequencing (PhIP-Seq). Here, an oligonucleotide library encoding 56-mer peptides across the proteomes of all known pathogenic human viruses (~400 species and strains) identified more than 800 highly selective epitopes in the SARS-CoV-2 proteome by machine learning.[142] Compared to these proteomewide analyses, high-sample-throughput serology assays using suspension bead arrays have been applied to home-sampled dried blood spots (DBS).[143] Altogether, these antibody-centric studies of the circulating proteome illustrated the systemic view of the immune response against human and pathogenic targets.

MICROSAMPLING OPPORTUNITIES

Blood-sampling techniques and preparatory processes have evolved to meet the requirements of individual clinical tests but have primarily not been optimized for use outside clinical settings. For cost-effective, offsite monitoring of subjects, the ideal sample should be collected quickly and robustly, with minimal involvement of clinical personnel, and easily stored while containing a wide array of biomolecules responsive to health and disease states. Recently, microsampling has reemerged to collect minimal invasive sample volumes, such as blood drops or interstitial fluid. Such samples can contain valuable biological information from proteins, metabolites, and lipids.[144−147]

These approaches offer advantages, including not requiring a phlebotomist or clinician and opening new avenues for understanding disease phenotypes through repeated longitudinal sampling. Patient-centric sampling methods can facilitate population-scale health assessments and high-frequency sampling to capture time-resolved changes in molecular phenotypes, such as during exercise or drug trials. However, the robustness of this sampling method across larger cohorts and the resulting analytical insights remain to be explored and confirmed.

Volumetric microsampling devices have emerged as alternatives to traditional dried blood spot (DBS) sampling, primarily due to their capability to collect precise and consistent sample volumes (e.g., 10, 20, or 30 μL). These devices also address the inherent challenge of volumetric hematocrit effects associated with DBS sampling. Alternative setups that filter blood cells recently emerged but have not been tested and proven to improve the analytical performance compared to classical whole blood collections. Their user-friendly nature, precision, and convenient storage make them a popular and cost-efficient alternative in population-scaled research and clinical studies.[143,148]

However, there are limitations when applying deep proteomics analysis to samples collected using these microsampling devices. Tools and techniques optimized for plasma proteomics may not yield the best results when applied to these samples. Alternatively, optimizations specific to microsampling samples may be required. For instance, blood contains large amounts of erythrocytes that further challenge the limited dynamic range by masking analytes from readout detectors. Drying proteins can produce stable salt crystals that remain difficult to resolve. Eluting protein from such matrices may require detergents or chemical treatments that affect the suitability of the samples for proteomics assays. Moreover, the increased sample heterogeneity comprising soluble proteins, vesicles, and cells requires adjustments in protein extraction protocols.

CONCLUSION AND OUTLOOK

The analysis of circulating proteins has gained growing interest, and several commercial providers (Table 1) have lowered the threshold to access robust technologies beyond the proteomics community. Historically, serum and plasma proteomics workflows have emerged from academic laboratories within the proteomics community. Given the significant promise of robust, deep, and accessible plasma proteome profiling, commercial platforms have recently emerged focusing on MS and non-MS readouts.[149] It remains to be seen how the established and more recent platforms will

https://doi.org/10.1021/acs.jproteome.4c00586
J. Proteome Res. XXXX, XXX, XXX−XXX

Table 2. Examples of Recent Protein Biomarker Projects Using Different Proteomics Techniques across Disease Areas

year	title of study[reference]	description of biomarker(s)	disease areas
2017	Role of Exosomal Proteins in Cancer Diagnosis[161]	several for different cancers	cancer
2020	Multiomic Blood Correlates of Genetic Risk Identify Presymptomatic Disease Alterations[162]	766 detectable alterations in proteomic, metabolomic, and standard clinical laboratory measurements	multiple
2020	Proteome Profiling in Cerebrospinal Fluid Reveals Novel Biomarkers of Alzheimer'S Disease[163]	26 protein panel	neurological
2021	A Time-Resolved Proteomic and Prognostic Map of COVID-19[164]	different panels for different outcome predictions like altered coagulation and inflammation	infection
2021	Plasma Proteome Fingerprints Reveal Distinctiveness and Clinical Outcome of SARS-CoV-2 Infection[165]	ADM, IL-6, MCP-3, TRAIL-R2, and PD-L1 predictive for death	infection
2021	Lung Proteomic Biomarkers Associated with Chronic Obstructive Pulmonary Disease[166]	25 proteins associated with COPD	infection
2022	Noninvasive Proteomic Biomarkers for Alcohol-Related Liver Disease[109]	nine protein panel for fibrosis, 12 protein panel for steatosis	metabolism
2023	Next Generation Pan-Cancer Blood Proteome Profiling Using Proximity Extension Assay[167]	different proteins and panels for specific cancer types	cancer
2023	Targeted Plasma Proteomics Reveals Signatures Discriminating COVID-19 from Sepsis with Pneumonia[168]	TRIM21, PTN, and CASP8	infection
2023	Elevated Plasma Complement Factor H Related 5 Protein Is Associated with Venous Thromboembolism[20]	CFHR5	cardiovascular
2023	Associations of Plasma Proteomics with Type 2 Diabetes and Related Traits: Results from the Longitudinal KORA S4/F4/FF4 Study[169]	protein panels for subtypes of prevalent prediabetes	metabolism
2023	Mass Spectrometry-Based Autoimmune Profiling Reveals Predictive Autoantigens in Idiopathic Pulmonary Fibrosis[170]	immunoglobulin profiling in plasma reveals autoimmune signature	respiratory
2023	Proteomics Reveal Biomarkers for Diagnosis, Disease Activity and Long-Term Disability Outcomes in Multiple Sclerosis[171]	11 proteins in CSF, none in plasma	neurological
2022	Stratifin as a Novel Diagnostic Biomarker in Serum For Diffuse Alveolar Damage[172]	eight proteins from SOMA scan are validated	respiratory
2023	Proteome Profiling of Early Gestational Plasma Reveals Novel Biomarkers of Congenital Heart Disease[173]	nine proteins predict CHD with high accuracy	cardiovascular
2022	Circulating Proteomic Panels for Risk Stratification of Intracranial Aneurysm and Its Rupture[174]	two sets of biomarker combinations to accurately distinguish IA from healthy controls (accuracy: 87.50%) or classify IA rupture patients (accuracy: 91.67%)	cardiovascular
2015	Prediction of Colorectal Cancer Diagnosis Based on Circulating Plasma Proteins[175]	five proteins as predictive diagnostic signature	cancer
2020	Extracellular Vesicle and Particle Biomarkers Define Multiple Human Cancers[176]	defined a panel of tumor-type-specific EVP proteins in TEs and plasma, which can classify tumors of unknown primary origin	cancer
2023	Absolute Quantification of Pan-Cancer Plasma Proteomes Reveals Unique Signature in Multiple Myeloma[177]	potential biomarker panel of seven protein targets for the diagnosis of multiple myeloma patients	cancer
2023	High-Throughput Plasma Proteomics to Define the Precursor Multiple Myeloma Proteome and Identify Candidate High-Risk Disease Biomarkers of Progression[178]	identified proteins that significantly distinguished monoclonal gammopathy of undetermined significance (MGUS), smoldering multiple myeloma (SMM), and NDMM from healthy donors	cancer
2023	Plasma Proteomic Associations with Genetics and Health in the UK Biobank[68]	protein quantitative trait locus (pQTL) mapping of 2923 proteins that identifies 14 287 primary genetic associations, of which 81% are previously undescribed	multiple
2023	Large-Scale Plasma Proteomics Comparisons through Genetics and Disease Associations[22]	Olink Explore 3072 data generated by the UK Biobank Pharma Proteomics Project1 on plasma samples from more than 50 000 UK Biobank participants and compared the results with those of a SomaScan v4 study on plasma from 36 000 Icelandic people2, for 1514 of whom Olink data were also available	multiple
2023	Rare Variant Associations with Plasma Protein Levels in the UK Biobank[67]	variant-level exomewide association study identified 5433 rare genotype−protein associations, of which 81% were undetected in a previous genomewide association study of the same cohort	multiple
2024	High Throughput Plasma Proteomics and Risk of Heart Failure and Frailty in Late Life[179]	18 individual proteins with consistent associations with risk of incident heart failure in mid- and late life	heart failure and frailty

https://doi.org/10.1021/acs.jproteome.4c00586
J. Proteome Res. XXXX, XXX, XXX−XXX

complement each other to improve our understanding of circulating proteins (and besides being faster, cheaper, more sensitive, of higher throughput, and requiring low sample volumes).

Studying the circulating proteome is still an ideal showcase for technology-driven biology. We have and will continue to learn about personalized health signatures and how these can change during our lives due to the development of diseases. We will better understand the (many) processes in which pleiotropic proteins are involved and which sample, method, or phenotype-related traits influence protein detection and analysis. We will likely have many computational tools available to capture quality features and reveal a diversity of underlying phenotypes (e.g., use of medication, liver metabolism, or kidney damage) rather than sticking to the silos of categories of being grouped as healthy or diseased. The ease of access to blood will also allow us to use new sampling schemes to determine systemwide molecular health trajectories, such as organ-specific aging,[159] at a much higher frequency than currently possible with clinical sampling and the available large biobanks. The hits and misses of blood proteome studies will likely influence how we study other but connected body fluids such as CSF, urine, or saliva.

In the coming years, we can expect an even more noticeable proportion of the circulating proteins to be reliably measured, ideally by several different technology platforms. While the growing adaptation, depth, throughput, and speed will be a natural development and find its use in a growing research community with interests in different types of biomarker studies (Table 2), it remains to be proven how these technological advances can add to biology and translate into tests that aid to improve or even save lives. New tools will be needed to test observations from blood analyses in functional and systemwide studies, where non-modifiable factors (e.g., genetic variation) and perturbations (e.g., lifestyle, infectious agents, or medication) can be decoded to gain mechanistic insights into the human 'omes.[160] In addition, the description of the circulating proteome will become more complete as soon as we can reliably add additional layers of complemtary information on top of protein levels, determine the protein interaction landscape in blood, capture the diversity of protein modifications and structural elements, and detect the array of single-molecule variants.

As technological progress continues and proteomics data become increasingly utilized by researchers outside the proteomics community, we must remain transparent and primarily report how many proteins we could measure in all samples (and not the sum of all detected). We need to be aware of the increasing request for platform-independent reliability in identifying circulating proteins, especially for the newer entries, those of the lowest abundance, and those prone to be influenced by molecular and analytical variance.

- Mass spectrometry and affinity-based assays have enhanced the utility of studying the circulating proteome.
- Persisting challenges emphasize the necessity to ensure data quality and reliability to disentangle biology.
- Undervaluing the importance of how confounding factors hamper the sample integrity can decrease the potential to reproduce findings.
- Future perspectives advocate for a broader approach considering extracellular vesicles, additional categories, and relationships between circulating proteins.
- Alternative sampling strategies can be implemented to comprehend the dynamic of human health and disease.

Pending issues include the following:

- Providing evidence for the presence of all 20 000 proteins in circulation.
- Detecting all circulating proteins in all samples studied.
- Quantifying the concentration of circulating proteins across the abundance ranges.
- Cataloging the PTM and interaction landscape of circulating proteins.
- Defining quality and integrity criteria for biological samples.
- Implementing practices for preanalytical, interindividual, and genetic factors.
- Identifying the sources of unknown biases and variation.
- Annotating the proteoforms and their regulation in health and disease.
- Implementing novel sampling techniques for deep proteome analysis.

■ ASSOCIATED CONTENT

Data Availability Statement

Data are available at https://peptideatlas.org/builds/human/plasma/

■ AUTHOR INFORMATION

Corresponding Authors

Philipp E. Geyer − *Department of Proteomics and Signal Transduction, Max Planck Institute of Biochemistry, 82152 Martinsried, Germany;* orcid.org/0000-0001-7980-4826; Email: pgeyer@biochem.mpg.de

Eric W. Deutsch − *Institute for Systems Biology, Seattle, Washington 98109, United States;* orcid.org/0000-0001-8732-0928; Email: edeutsch@systemsbiology.org

Jochen M. Schwenk − *Science for Life Laboratory, Department of Protein Science, KTH Royal Institute of Technology, 17121 Solna, Sweden;* orcid.org/0000-0001-8141-8449; Email: jochen.schwenk@scilifelab.se

Authors

Daniel Hornburg − *Seer, Inc., Redwood City, California 94065, United States; Bruker Scientific, San Jose, California 95134, United States;* orcid.org/0000-0002-6618-7774

Maria Pernemalm − *Department of Oncology and Pathology/Science for Life Laboratory, Karolinska Institutet, 17177 Stockholm, Sweden*

Stefanie M. Hauck − *Metabolomics and Proteomics Core, Helmholtz Zentrum München GmbH, German Research Center for Environmental Health, 85764 Oberschleissheim, Munich, Germany;* orcid.org/0000-0002-1630-6827

Krishnan K. Palaniappan − *Freenome, South San Francisco, California 94080, United States;* orcid.org/0000-0003-0203-9016

Vincent Albrecht − *Department of Proteomics and Signal Transduction, Max Planck Institute of Biochemistry, 82152 Martinsried, Germany*

Laura F. Dagley − *The Walter and Eliza Hall Institute for Medical Research, Parkville, VIC 3052, Australia;*

https://doi.org/10.1021/acs.jproteome.4c00586
J. Proteome Res. XXXX, XXX, XXX−XXX

Department of Medical Biology, University of Melbourne, Parkville, VIC 3052, Australia; orcid.org/0000-0003-4171-3712

Robert L. Moritz − *Institute for Systems Biology, Seattle, Washington 98109, United States*

Xiaobo Yu − *State Key Laboratory of Medical Proteomics, Beijing Proteome Research Center, National Center for Protein Sciences-Beijing (PHOENIX Center), Beijing Institute of Lifeomics, Beijing 102206, China*

Fredrik Edfors − *Science for Life Laboratory, Department of Protein Science, KTH Royal Institute of Technology, 17121 Solna, Sweden*

Yves Vandenbrouck − *CEA, Fundamental Research Division, 91191 Gif-sur-Yvette, France;* orcid.org/0000-0002-1292-373X

Johannes B. Mueller-Reif − *Department of Proteomics and Signal Transduction, Max Planck Institute of Biochemistry, 82152 Martinsried, Germany*

Zhi Sun − *Institute for Systems Biology, Seattle, Washington 98109, United States*

Virginie Brun − *Université Grenoble Alpes, CEA, Leti, Clinatec, Inserm UA13 BGE, CNRS, FR2048 Grenoble, France;* orcid.org/0000-0002-9098-8707

Sara Ahadi − *Alkahest, Inc., Suite D San Carlos, California 94070, United States*

Gilbert S. Omenn − *Institute for Systems Biology, Seattle, Washington 98109, United States; Departments of Computational Medicine & Bioinformatics, Internal Medicine, Human Genetics and Environmental Health, University of Michigan, Ann Arbor, Michigan 48109-2218, United States;* orcid.org/0000-0002-8976-6074

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jproteome.4c00586

Funding

Open access funded by Max Planck Society. This work was funded in part by the National Institutes of Health grants R01 GM087221 (E.W.D., R.L.M.) and U19 AG023122 (R.L.M.), by the National Science Foundation grants DBI-2324882 (E.W.D.) and DBI-1933311 (E.W.D.), and by the German Ministry for Science and Education funding action CLINSPECT-M [FKZ161L0214E] to S.M.H. G.S.O. acknowledges support from National Institutes of Health Grants P30ES017885-01A1 and U24CA210967. J.M.S. acknowledges the Knut and Alice Wallenberg Foundation's support for the Human Protein Atlas (HPA).

Notes

The authors declare the following competing financial interest(s): At the time of the first submission, D.H. was an employee of Bruker Scientific, K.K.P. was an employee of Freenome, and S.A. was an employee of Alkahest. The other authors declare that they have no conflict of interest.

■ ACKNOWLEDGMENTS

We thank the community for providing plasma proteomics data and recognize that we might have missed many emerging concepts and published studies.

■ GLOSSARY

antibody depletion, this technology removes proteins with a high concentration typically from plasma samples, allowing the detection of lower concentrated proteins by MS-based analysis

assays, tests or methods used to detect proteins

(auto)antibody isotyping, identifying different types and classes of antibodies in the body, which are crucial components of the immune system; this analysis provides insights into immune responses and disease mechanisms

biological variation, nontechnical differences in protein measurements and other biological molecules among individuals

biomarker, a biological factor found in blood, tissues, or other body fluids that can indicate a normal or abnormal process, a condition, or a disease; biomarkers are used to diagnose diseases, monitor diseases and response to therapy, and predict disease outcomes

biomarker discovery, the process of finding molecules in biological samples that can indicate the presence of a disease or health condition

circulating proteome, the collection of all proteins found in the liquid part of blood; scientists study these proteins to understand health and disease

confounding factors, external parameters that can affect study results, which need to be considered

dynamic range, the range between the proteins with the highest and the lowest concentration

extracellular vesicles, lipid particles released by cells that carry proteins and other molecules

false discovery rate, the percentage of incorrect identifications caused by performing a larger number of tests; a lower false discovery rate indicates more accurate results

glycoproteomics, the technology to study glycosylated proteins or peptides

liquid handling, these are automated robotic systems used in laboratories to process and analyze biological samples, such as blood or plasma, with high precision

mass spectrometry (MS) and affinity-based approaches, advanced techniques used by scientists to measure and analyze proteins in blood samples

microsampling, collecting tiny amounts of blood for analysis

peptides, short chains of amino acids, the building blocks of proteins; peptides are a preparatory product of proteins and play a role in protein analysis

plasma, a type of cell-free blood preparation that uses additives to prevent coagulation from occurring

post-translational modifications, changes or modifications of the molecular composition of proteins after they are made, affecting their structure, functions, and interaction partners

preanalytics, processes required to collect and prepare samples for analysis, like blood collection, storage, or shipment

protein abundance, the amount of a specific protein present in a sample; scientists measure protein abundance to understand the concentration levels of different proteins in the blood

protein interaction, how proteins interact with each other in the body, affecting various biological processes; scientists study these interactions to understand the functions of proteins

proteomics, the study of all the proteins in a cell, tissue, or organism; in this context, it refers to understanding the proteins present in blood

proteoform, versions or modifications of proteins

https://doi.org/10.1021/acs.jproteome.4c00586
J. Proteome Res. XXXX, XXX, XXX−XXX

sample contamination, unwanted substances enter the blood sample and affect the outcome of test results
sample preparation, methods and techniques used to prepare biological samples for analysis
serum, a type of cell-free blood preparation where coagulation occurs and clotting factors precipitate
targeted proteomics, this technique allows scientists to focus on specific and preselected proteins or peptides in a sample, helping to detect and quantify them accurately.

REFERENCES

(1) Anderson, N. L.; Anderson, N. G. The Human Plasma Proteome. *Mol. Cell. Proteomics* **2002**, *1*, 845−867.
(2) Deutsch, E. W.; et al. Advances and Utility of the Human Plasma Proteome. *J. Proteome Res.* **2021**, *20*, 5241−5263.
(3) Ignjatovic, V.; et al. Mass Spectrometry-Based Plasma Proteomics: Considerations from Sample Collection to Achieving Translational Data. *J. Proteome Res.* **2019**, *18*, 4085−4097.
(4) Omenn, G. S.; et al. Overview of the HUPO Plasma Proteome Project: results from the pilot phase with 35 collaborating laboratories and multiple analytical groups, generating a core dataset of 3020 proteins and a publicly-available database. *Proteomics* **2005**, *5*, 3226−3245.
(5) Dodig-Crnković, T.; et al. Facets of individual-specific health signatures determined from longitudinal plasma proteome profiling. *EBioMedicine* **2020**, *57*, No. 102854.
(6) Geyer, P. E.; et al. Proteomics reveals the effects of sustained weight loss on the human plasma proteome. *Mol. Syst. Biol.* **2016**, *12*, 901.
(7) Bader, J. M.; Albrecht, V.; Mann, M. MS-Based Proteomics of Body Fluids: The End of the Beginning. *Mol. Cell. Proteomics* **2023**, *22*, 100577.
(8) Henry, A.; et al. Therapeutic Targets for Heart Failure Identified Using Proteomics and Mendelian Randomization. *Circulation* **2022**, *145*, 1205−1217.
(9) Suhre, K.; McCarthy, M. I.; Schwenk, J. M. Genetics meets proteomics: perspectives for large population-based studies. *Nat. Rev. Genet.* **2021**, *22*, 19−37.
(10) Assarsson, E.; et al. Homogenous 96-plex PEA immunoassay exhibiting high sensitivity, specificity, and excellent scalability. *PloS One* **2014**, *9*, No. e95192.
(11) Gold, L.; et al. Aptamer-based multiplexed proteomic technology for biomarker discovery. *PloS One* **2010**, *5*, No. e15004.
(12) Ferkingstad, E.; et al. Large-scale integration of the plasma proteome with genetics and disease. *Nat. Genet.* **2021**, *53*, 1712−1721.
(13) Folkersen, L.; et al. Genomic and drug target evaluation of 90 cardiovascular proteins in 30,931 individuals. *Nat. Metab.* **2020**, *2*, 1135−1148.
(14) Sun, B. B.; et al. Genomic atlas of the human plasma proteome. *Nature* **2018**, *558*, 73−79.
(15) Zhong, W.; et al. Next generation plasma proteome profiling to monitor health and disease. *Nat. Commun.* **2021**, *12*, 2493.
(16) Feng, W.; Beer, J. C.; Hao, Q.; Ariyapala, I. S.; Sahajan, A.; Komarov, A.; Cha, K.; Moua, M.; Qiu, X.; Xu, X.; Iyengar, S.; Yoshimura, T.; Nagaraj, R.; Wang, L.; Yu, M.; Engel, K.; Zhen, L.; Xue, W.; Lee, C.-j.; Park, C. H.; Peng, C.; Zhang, K.; Grzybowski, A.; Hahm, J.; Schmidt, S. V.; Odainic, A.; Spitzer, J.; Buddika, K.; Kuo, D.; Fang, L.; Zhang, B.; Chen, S.; Latz, E.; Yin, Y.; Luo, Y.; Ma, X.-J. NULISA: A novel proteomic liquid biopsy platform with attomolar sensitivity and high multiplexing. *Nat. Comm.* **2023**, *14*, 7238.
(17) Koprulu, M.; et al. Proteogenomic links to human metabolic diseases. *Nat. Metab.* **2023**, *5*, 516−528.
(18) Katz, D. H.; et al. Proteomic profiling platforms head to head: Leveraging genetics and clinical traits to compare aptamer- and antibody-based methods. *Sci. Adv.* **2022**, *8*, No. eabm5164.
(19) Pietzner, M.; et al. Synergistic insights into human health from aptamer- and antibody-based proteomic profiling. *Nat. Commun.* **2021**, *12*, 6822.
(20) Iglesias, M. J.; et al. Elevated plasma complement factor H related 5 protein is associated with venous thromboembolism. *Nat. Commun.* **2023**, *14*, 3280.
(21) Petrera, A.; et al. Multiplatform Approach for Plasma Proteomics: Complementarity of Olink Proximity Extension Assay Technology to Mass Spectrometry-Based Protein Profiling. *J. Proteome Res.* **2021**, *20*, 751−762.
(22) Eldjarn, G. H.; et al. Large-scale plasma proteomics comparisons through genetics and disease associations. *Nature* **2023**, *622*, 348−358.
(23) Omenn, G. S.; et al. The 2022 Report on the Human Proteome from the HUPO Human Proteome Project. *J. Proteome Res.* **2023**, *22*, 1024−1042.
(24) Guzman, U. H.; et al. Ultra-fast label-free quantification and comprehensive proteome coverage with narrow-window data-independent acquisition. *Nat. Biotechnol.* **2024**, 1−12.
(25) Heil, L. R.; et al. Evaluating the Performance of the Astral Mass Analyzer for Quantitative Proteomics Using Data-Independent Acquisition. *J. Proteome Res.* **2023**, *22*, 3290−3300.
(26) Hendricks, N. G.; et al. An Inflection Point in High-Throughput Proteomics with Orbitrap Astral: Analysis of Biofluids, Cells, and Tissues. *J. Proteome Res.* **2024**, *23*, 4163−4169.
(27) Geyer.; et al. Plasma Proteome Profiling to Assess Human Health and Disease. *Cell Syst.* **2016**, *2*, 185−195.
(28) Messner, C. B.; et al. Ultra-High-Throughput Clinical Proteomics Reveals Classifiers of COVID-19 Infection. *Cell Syst.* **2020**, *11*, 11−24.e4.
(29) King, C. D.; Kapp, K. L.; Arul, A. B.; Choi, M. J.; Robinson, R. A. S. Advancements in automation for plasma proteomics sample preparation. *Mol. Omics* **2022**, *18*, 828−839.
(30) Fu, Q.; Murray, C. I.; Karpov, O. A.; Van Eyk, J. E. Automated proteomic sample preparation: The key component for high throughput and quantitative mass spectrometry analysis. *Mass Spectrom. Rev.* **2023**, *42*, No. 873.
(31) Bache, N.; et al. A Novel LC System Embeds Analytes in Pre-formed Gradients for Rapid, Ultra-robust Proteomics. *Mol. Cell. Proteomics MCP* **2018**, *17*, 2284−2296.
(32) Bian, Y.; et al. Robust Microflow LC-MS/MS for Proteome Analysis: 38 000 Runs and Counting. *Anal. Chem.* **2021**, *93*, 3686−3690.
(33) Messner, C. B.; et al. Ultra-fast proteomics with Scanning SWATH. *Nat. Biotechnol.* **2021**, *39*, 846−854.
(34) Geyer, P. E.; et al. High-resolution serum proteome trajectories in COVID-19 reveal patient-specific seroconversion. *EMBO Mol. Med.* **2021**, *13*, No. e14167.
(35) Koh, B.; et al. Multi-omics profiling with untargeted proteomics for blood-based early detection of lung cancer. 2024; medRxiv preprint at DOI: 10.1101/2024.01.03.24300798.
(36) Viode, A.; et al. A simple, time- and cost-effective, high-throughput depletion strategy for deep plasma proteomics. *Sci. Adv.* **2023**, *9*, No. eadf9717.
(37) Mann, M.; Kumar, C.; Zeng, W.-F.; Strauss, M. T. Artificial intelligence for proteomics and biomarker discovery. *Cell Syst.* **2021**, *12*, 759−770.
(38) Torun, F. M.; et al. Transparent Exploration of Machine Learning for Biomarker Discovery from Proteomics and Omics Data. *J. Proteome Res.* **2023**, *22*, 359−367.
(39) Tognetti, M.; et al. Biomarker Candidates for Tumors Identified from Deep-Profiled Plasma Stem Predominantly from the Low Abundant Area. *J. Proteome Res.* **2022**, *21*, 1718−1735.
(40) Pieper, R.; et al. Multi-component immunoaffinity subtraction chromatography: an innovative step towards a comprehensive survey of the human plasma proteome. *Proteomics* **2003**, *3*, 422−432.
(41) Keshishian, H.; et al. Quantitative, multiplexed workflow for deep analysis of human blood plasma and biomarker discovery by mass spectrometry. *Nat. Protoc.* **2017**, *12*, 1683−1701.

https://doi.org/10.1021/acs.jproteome.4c00586
J. Proteome Res. XXXX, XXX, XXX−XXX

(42) Ferdosi, S.; et al. Enhanced Competition at the Nano-Bio Interface Enables Comprehensive Characterization of Protein Corona Dynamics and Deep Coverage of Proteomes. *Adv. Mater.* **2022**, *34*, No. e2206008.
(43) Ferdosi, S.; et al. Engineered nanoparticles enable deep proteomics studies at scale by leveraging tunable nano-bio interactions. *Proc. Natl. Acad. Sci. U. S. A.* **2022**, *119*, No. e2106053119.
(44) Suhre, K.; Venkataraman, G. R.; Guturu, H.; Halama, A.; Stephan, N.; Thareja, G.; Sarwath, H.; Motamedchaboki, K.; Donovan, M. K. R.; Siddiqui, A.; Batzoglou, S.; Schmidt, F. Nanoparticle Enrichment Mass-Spectrometry Proteomics Identifies Protein Altering Variants for Precise pQTL Mapping. *Nat. Comm.* **2024**, *15*, 989.
(45) Viode, A.; et al. Longitudinal plasma proteomic analysis of 1117 hospitalized patients with COVID-19 identifies features associated with severity and outcomes. *Sci. Adv.* **2024**, *10*, No. eadl5762.
(46) Geyer, P. E.; et al. Plasma Proteome Profiling to detect and avoid sample-related biases in biomarker studies. *EMBO Mol. Med.* **2019**, *11*, No. e10427.
(47) Marx, V. Targeted proteomics. *Nat. Methods* **2013**, *10*, 19−22.
(48) Brzhozovskiy, A.; et al. The Parallel Reaction Monitoring-Parallel Accumulation-Serial Fragmentation (prm-PASEF) Approach for Multiplexed Absolute Quantitation of Proteins in Human Plasma. *Anal. Chem.* **2022**, *94*, 2016−2022.
(49) Lesur, A.; et al. Quantification of 782 Plasma Peptides by Multiplexed Targeted Proteomics. *J. Proteome Res.* **2023**, *22*, 1630−1638.
(50) Wang, Z.; et al. A multiplex protein panel assay for severity prediction and outcome prognosis in patients with COVID-19: An observational multi-cohort study. *eClinicalMedicine* **2022**, *49*, No. 101495.
(51) Demichev, V.; et al. A proteomic survival predictor for COVID-19 patients in intensive care. *PLOS Digit. Health* **2022**, *1*, No. e0000007.
(52) Captur, G.; et al. Plasma proteomic signature predicts who will get persistent symptoms following SARS-CoV-2 infection. *eBioMedicine* **2022**, *85*, No. 104293.
(53) Wahle, M.; et al. A novel hybrid high speed mass spectrometer allows rapid translation from biomarker candidates to targeted clinical tests using 15N labeled proteins. 2024; bioRxiv preprint at DOI: 10.1101/2024.06.02.597029.
(54) Gallien, S.; Kim, S. Y.; Domon, B. Large-Scale Targeted Proteomics Using Internal Standard Triggered-Parallel Reaction Monitoring (IS-PRM). *Mol. Cell. Proteomics MCP* **2015**, *14*, 1630−1644.
(55) Grossegesse, M.; Hartkopf, F.; Nitsche, A.; Doellinger, J. Stable Isotope-Triggered Offset Fragmentation Allows Massively Multiplexed Target Profiling on Quadrupole-Orbitrap Mass Spectrometers. *J. Proteome Res.* **2020**, *19*, 2854−2862.
(56) Erickson, B. K.; et al. A Strategy to Combine Sample Multiplexing with Targeted Proteomics Assays for High-Throughput Protein Signature Characterization. *Mol. Cell* **2017**, *65*, 361−370.
(57) Rougemont, B.; et al. Scout-MRM: Multiplexed Targeted Mass Spectrometry-Based Assay without Retention Time Scheduling Exemplified by Dickeya dadantii Proteomic Analysis during Plant Infection. *Anal. Chem.* **2017**, *89*, 1421−1426.
(58) Wichmann, C.; et al. MaxQuant.Live Enables Global Targeting of More Than 25,000 Peptides. *Mol. Cell. Proteomics* **2019**, *18*, 982−994.
(59) Oeckl, P.; Steinacker, P.; Otto, M. Comparison of Internal Standard Approaches for SRM Analysis of Alpha-Synuclein in Cerebrospinal Fluid. *J. Proteome Res.* **2018**, *17*, 516−523.
(60) Hober, A.; et al. Absolute Quantification of Apolipoproteins Following Treatment with Omega-3 Carboxylic Acids and Fenofibrate Using a High Precision Stable Isotope-labeled Recombinant Protein Fragments Based SRM Assay. *Mol. Cell. Proteomics MCP* **2019**, *18*, 2433−2446.
(61) Zeiler, M.; Straube, W. L.; Lundberg, E.; Uhlen, M.; Mann, M. A Protein Epitope Signature Tag (PrEST) Library Allows SILAC-based Absolute Quantification and Multiplexed Determination of Protein Copy Numbers in Cell Lines. *Mol. Cell. Proteomics* **2012**, *11*, No. O111.009613.
(62) Deutsch, E. W.; et al. The ProteomeXchange consortium in 2017: supporting the cultural change in proteomics public data deposition. *Nucleic Acids Res.* **2017**, *45*, D1100−D1106.
(63) Deutsch, E. W.; et al. The ProteomeXchange consortium at 10 years: 2023 update. *Nucleic Acids Res.* **2023**, *51*, D1539−D1548.
(64) Melani, R. D.; et al. The Blood Proteoform Atlas: A reference map of proteoforms in human hematopoietic cells. *Science* **2022**, *375*, 411−418.
(65) Vizcaíno, J. A.; et al. ProteomeXchange provides globally coordinated proteomics data submission and dissemination. *Nat. Biotechnol.* **2014**, *32*, 223−226.
(66) Choi, M.; et al. MassIVE.quant: a community resource of quantitative mass spectrometry-based proteomics datasets. *Nat. Methods* **2020**, *17*, 981−984.
(67) Dhindsa, R. S.; et al. Rare variant associations with plasma protein levels in the UK Biobank. *Nature* **2023**, *622*, 339−347.
(68) Sun, B. B.; et al. Plasma proteomic associations with genetics and health in the UK Biobank. *Nature* **2023**, *622*, 329−338.
(69) PRIDE-Archive/pride-resources. *GitHub*; https://github.com/PRIDE-Archive/pride-resources/blob/f911140923a708e19c78bc664116b9f5481a5b03/guidelines/pride-affinity-proteomics.md.
(70) Beutgen, V. M.; et al. Secretome Analysis Using Affinity Proteomics and Immunoassays: A Focus on Tumor Biology. *Mol. Cell. Proteomics* **2024**, *23*, No. 100830.
(71) Kurgan, N.; Kjærgaard Larsen, J.; Deshmukh, A. S. Harnessing the power of proteomics in precision diabetes medicine. *Diabetologia* **2024**, *67*, 783−797.
(72) Farrah, T.; et al. A high-confidence human plasma proteome reference set with estimated concentrations in PeptideAtlas. *Mol. Cell. Proteomics* **2011**, *10*, No. M110.006353.
(73) Schwenk, J. M.; et al. The Human Plasma Proteome Draft of 2017: Building on the Human Plasma PeptideAtlas from Mass Spectrometry and Complementary Assays. *J. Proteome Res.* **2017**, *16*, 4299−4310.
(74) Deutsch, E. W.; et al. Human Plasma PeptideAtlas. *Proteomics* **2005**, *5*, 3497−3500.
(75) Deutsch, E. W.; et al. Trans-Proteomic Pipeline, a standardized data processing pipeline for large-scale reproducible proteomics informatics. *PROTEOMICS - Clin. Appl.* **2015**, *9*, 745−754.
(76) Deutsch, E. W.; et al. Human Proteome Project Mass Spectrometry Data Interpretation Guidelines 3.0. *J. Proteome Res.* **2019**, *18*, 4108−4116.
(77) Zahn-Zabal, M.; et al. The neXtProt knowledgebase in 2020: data, tools and usability improvements. *Nucleic Acids Res.* **2019**, *48*, D328−D334.
(78) Ramazi, S.; Zahiri, J. Post-translational modifications in proteins: resources, tools and prediction methods. *Database J. Biol. Databases Curation* **2021**, *2021*, No. baab012.
(79) Leutert, M.; Entwisle, S. W.; Villén, J. Decoding Post-Translational Modification Crosstalk With Proteomics. *Mol. Cell. Proteomics MCP* **2021**, *20*, No. 100129.
(80) Venne, A. S.; Kollipara, L.; Zahedi, R. P. The next level of complexity: crosstalk of posttranslational modifications. *Proteomics* **2014**, *14*, 513−524.
(81) Reily, C.; Stewart, T. J.; Renfrow, M. B.; Novak, J. Glycosylation in health and disease. *Nat. Rev. Nephrol.* **2019**, *15*, 346−366.
(82) Palaniappan, K. K.; Bertozzi, C. R. Chemical Glycoproteomics. *Chem. Rev.* **2016**, *116*, 14277−14306.
(83) Suttapitugsakul, S.; Sun, F.; Wu, R. Recent Advances in Glycoproteomic Analysis by Mass Spectrometry. *Anal. Chem.* **2020**, *92*, 267−291.

https://doi.org/10.1021/acs.jproteome.4c00586
J. Proteome Res. XXXX, XXX, XXX−XXX

(84) Kawahara, R.; et al. Community evaluation of glycoproteomics informatics solutions reveals high-performance search strategies for serum glycopeptide analysis. *Nat. Methods* **2021**, *18*, 1304−1316.
(85) Kong, S.; et al. pGlycoQuant with a deep residual network for quantitative glycoproteomics at intact glycopeptide level. *Nat. Commun.* **2022**, *13*, 7539.
(86) Drake, P. M.; et al. Sweetening the pot: adding glycosylation to the biomarker discovery equation. *Clin. Chem.* **2010**, *56*, 223−236.
(87) Felder, M.; et al. MUC16 (CA125): tumor biomarker to cancer therapy, a work in progress. *Mol. Cancer* **2014**, *13*, 129.
(88) Wang, M.; Zhu, J.; Lubman, D. M.; Gao, C. Aberrant glycosylation and cancer biomarker discovery: a promising and thorny journey. *Clin. Chem. Lab. Med.* **2019**, *57*, 407−416.
(89) Ideo, H.; et al. Study of glycosylation of prostate-specific antigen secreted by cancer tissue-originated spheroids reveals new candidates for prostate cancer detection. *Sci. Rep.* **2020**, *10*, 2708.
(90) Pont, L.; et al. Site-Specific N-Linked Glycosylation Analysis of Human Carcinoembryonic Antigen by Sheathless Capillary Electrophoresis−Tandem Mass Spectrometry. *J. Proteome Res.* **2021**, *20*, 1666−1675.
(91) Chernykh, A.; Kawahara, R.; Thaysen-Andersen, M. Towards structure-focused glycoproteomics. *Biochem. Soc. Trans.* **2021**, *49*, 161−186.
(92) Wessels, H. J.; Kulkarni, P.; van Dael, M.; Suppers, A.; Willems, E.; Zijlstra, F.; Kragt, E.; Gloerich, J.; Schmit, P.-O.; Pengelley, S.; Marx, K.; van Gool, A. J. Plasma glycoproteomics delivers high-specificity disease biomarkers by detecting site-specific glycosylation abnormalities. *J. Adv. Res.* **2024**, *61*, 179−192.
(93) Pan, J.; et al. Glycoproteomics-based signatures for tumor subtyping and clinical outcome prediction of high-grade serous ovarian cancer. *Nat. Commun.* **2020**, *11*, 6139.
(94) Cao, L.; et al. Proteogenomic Characterization of Pancreatic Ductal Adenocarcinoma. *Cell* **2021**, *184*, 5031−5052.e26.
(95) Li, Q. K.; et al. Proteomic characterization of primary and metastatic prostate cancer reveals reduced proteinase activity in aggressive tumors. *Sci. Rep.* **2021**, *11*, 18936.
(96) Li, Y.; et al. Histopathologic and proteogenomic heterogeneity reveals features of clear cell renal cell carcinoma aggressiveness. *Cancer Cell* **2023**, *41*, 139−163.e17.
(97) Wang, Y.; Chen, H. Protein glycosylation alterations in hepatocellular carcinoma: function and clinical implications. *Oncogene* **2023**, *42*, 1970.
(98) Tagliabracci, V. S.; Pinna, L. A.; Dixon, J. E. Secreted protein kinases. *Trends Biochem. Sci.* **2013**, *38*, 121−130.
(99) Klement, E.; Medzihradszky, K. F. Extracellular Protein Phosphorylation, the Neglected Side of the Modification. *Mol. Cell. Proteomics MCP* **2017**, *16*, 1−7.
(100) Giorgianni, F.; Beranova-Giorgianni, S. Phosphoproteome Discovery in Human Biological Fluids. *Proteomes* **2016**, *4*, 37.
(101) Ghosh, R.; Ahmed, R.; Ahmed, H.; Chatterjee, B. P. Phosphorylated Proteins from Serum: A Promising Potential Diagnostic Biomarker of Cancer. *Int. J. Mol. Sci.* **2022**, *23*, 12359.
(102) Clark, D. J.; et al. Integrated Proteogenomic Characterization of Clear Cell Renal Cell Carcinoma. *Cell* **2019**, *179*, 964−983.e31.
(103) Lih, T. M.; et al. Detection of Pancreatic Ductal Adenocarcinoma-Associated Proteins in Serum. *Mol. Cell. Proteomics MCP* **2024**, *23*, No. 100687.
(104) Cao, L.; et al. Proteogenomic Characterization of Pancreatic Ductal Adenocarcinoma. *Cell* **2021**, *184*, 5031−5052.e26.
(105) Gillette, M. A.; et al. Proteogenomic characterization reveals therapeutic vulnerabilities in lung adenocarcinoma. *Cell* **2020**, *182*, 200−225.e35.
(106) Li, G. X. Comprehensive proteogenomic characterization of rare kidney tumors. *Cell Rep. Med.* **2023**, *5*, 3127.
(107) Filbin, M. R.; et al. Longitudinal proteomic analysis of severe COVID-19 reveals survival-associated signatures, tissue-specific cell death, and cell-cell interactions. *Cell Rep. Med.* **2021**, *2*, No. 100287.
(108) Pernemalm, M.; et al. In-depth human plasma proteome analysis captures tissue proteins and transfer of protein variants across the placenta. *eLife* **2019**, *8*, No. e41608.
(109) Niu, L.; et al. Noninvasive proteomic biomarkers for alcohol-related liver disease. *Nat. Med.* **2022**, *28*, 1277−1287.
(110) Cominetti, O.; et al. Proteomic Biomarker Discovery in 1000 Human Plasma Samples with Mass Spectrometry. *J. Proteome Res.* **2016**, *15*, 389−399.
(111) Dodig-Crnković, T.; Hong, M.-G.; Thomas, C. E.; Häussler, R. S.; Bendes, A.; Dale, M.; et al. Facets of Individual-Specific Health Signatures Determined from Longitudinal Plasma Proteome Profiling. *EBioMedicine* **2020**, *57*, 102854.
(112) Langfelder, P.; Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **2008**, *9*, 559.
(113) Kanonidis, E. I.; Roy, M. M.; Deighton, R. F.; Le Bihan, T. Protein Co-Expression Analysis as a Strategy to Complement a Standard Quantitative Proteomics Approach: Case of a Glioblastoma Multiforme Study. *PLoS One* **2016**, *11*, No. e0161828.
(114) Albrechtsen, N. J. W.; et al. Plasma Proteome Profiling Reveals Dynamics of Inflammatory and Lipid Homeostasis Markers after Roux-En-Y Gastric Bypass Surgery. *Cell Syst.* **2018**, *7*, 601−612.e3.
(115) Emilsson, V.; et al. Co-regulatory networks of human serum proteins link genetics to disease. *Science* **2018**, *361*, 769−773.
(116) Mi, Y.; et al. High-throughput mass spectrometry maps the sepsis plasma proteome and differences in response. *Sci. Trans. Med.* **2022**, *16*, 750.
(117) Demichev, V.; et al. A time-resolved proteomic and prognostic map of COVID-19. *Cell Syst.* **2021**, *12*, 780−794.e7.
(118) Price, N. D.; et al. A wellness study of 108 individuals using personal, dense, dynamic data clouds. *Nat. Biotechnol.* **2017**, *35*, 747−756.
(119) Tebani, A.; et al. Integration of molecular profiles in a longitudinal wellness profiling cohort. *Nat. Commun.* **2020**, *11*, 4487.
(120) Argelaguet, R.; et al. Multi-Omics Factor Analysis—a framework for unsupervised integration of multi-omics data sets. *Mol. Syst. Biol.* **2018**, *14*, No. e8124.
(121) Wu, C. C.; et al. Mag-Net: Rapid enrichment of membrane-bound particles enables high coverage quantitative analysis of the plasma proteome. 2023; bioRxiv preprint at DOI: 10.1101/2023.06.10.544439.
(122) Wu, D.; et al. Profiling surface proteins on individual exosomes using a proximity barcoding assay. *Nat. Commun.* **2019**, *10*, 3854.
(123) Karimi, N.; Dalirfardouei, R.; Dias, T.; Lötvall, J.; Lässer, C. Tetraspanins distinguish separate extracellular vesicle subpopulations in human serum and plasma − Contributions of platelet extracellular vesicles in plasma samples. *J. Extracell. Vesicles* **2022**, *11*, No. e12213.
(124) Muraoka, S.; et al. Comprehensive proteomic profiling of plasma and serum phosphatidylserine-positive extracellular vesicles reveals tissue-specific proteins. *iScience* **2022**, *25*, No. 104012.
(125) Vallejo, M. C.; et al. A proteomic meta-analysis refinement of plasma extracellular vesicles. *Sci. Data* **2023**, *10*, 837.
(126) Zhang, X.; Takeuchi, T.; Takeda, A.; Mochizuki, H.; Nagai, Y. Comparison of serum and plasma as a source of blood extracellular vesicles: Increased levels of platelet-derived particles in serum extracellular vesicle fractions alter content profiles from plasma extracellular vesicle fractions. *PLoS One* **2022**, *17*, No. e0270634.
(127) Hom, J. R.; Tomar, D.; Tipton, C. M. Exploring the Diversity of the B-Cell Receptor Repertoire Through High-Throughput Sequencing. In *Immune Receptors*; Rast, J., Buckley, K., Eds.; Springer US: New York, 2022; Vol. *2421*, pp 231−241.
(128) Dingess, K. A.; et al. Identification of common and distinct origins of human serum and breastmilk IgA1 by mass spectrometry-based clonal profiling. *Cell. Mol. Immunol.* **2023**, *20*, 26−37.
(129) Chen, J.; et al. Proteomic Analysis of Pemphigus Autoantibodies Indicates a Larger, More Diverse, and More Dynamic

https://doi.org/10.1021/acs.jproteome.4c00586
J. Proteome Res. XXXX, XXX, XXX−XXX

Repertoire than Determined by B Cell Genetics. *Cell Rep.* **2017**, *18*, 237−247.
(130) Bondt, A.; et al. Human plasma IgG1 repertoires are simple, unique, and dynamic. *Cell Syst.* **2021**, *12*, 1131−1143.e5.
(131) Bondt, A.; Dingess, K. A.; Hoek, M.; van Rijswijck, D. M. H.; Heck, A. J. R. A Direct MS-Based Approach to Profile Human Milk Secretory Immunoglobulin A (IgA1) Reveals Donor-Specific Clonal Repertoires With High Longitudinal Stability. *Front. Immunol.* **2021**, *12*, No. 789748.
(132) van Rijswijck, D. M. H.; et al. Discriminating cross-reactivity in polyclonal IgG1 responses against SARS-CoV-2 variants of concern. *Nat. Commun.* **2022**, *13*, 6103.
(133) Beaudoin-Bussières, G.; et al. Decline of Humoral Responses against SARS-CoV-2 Spike in Convalescent Individuals. *mBio* **2020**, *11*, No. e02590-20.
(134) Lv, H.; et al. Cross-reactive Antibody Response between SARS-CoV-2 and SARS-CoV Infections. *Cell Rep.* **2020**, *31*, No. 107725.
(135) Sterlin, D.; et al. IgA dominates the early neutralizing antibody response to SARS-CoV-2. *Sci. Transl. Med.* **2021**, *13*, No. eabd2223.
(136) Wang, Z.; et al. Enhanced SARS-CoV-2 neutralization by dimeric IgA. *Sci. Transl. Med.* **2021**, *13*, No. eabf1555.
(137) Larsen, M. D.; et al. Afucosylated IgG characterizes enveloped viral responses and correlates with COVID-19 severity. *Science* **2021**, *371*, No. eabc8378.
(138) Bournazos, S.; et al. Antibody fucosylation predicts disease severity in secondary dengue infection. *Science* **2021**, *372*, 1102−1105.
(139) Wang, D.; et al. AAgAtlas 1.0: a human autoantigen database. *Nucleic Acids Res.* **2017**, *45*, D769−D776.
(140) Wang, H.; et al. SARS-CoV-2 Proteome Microarray for Mapping COVID-19 Antibody Interactions at Amino Acid Resolution. *ACS Cent. Sci.* **2020**, *6*, 2238−2249.
(141) Jiang, H.; et al. SARS-CoV-2 proteome microarray for global profiling of COVID-19 specific IgG and IgM responses. *Nat. Commun.* **2020**, *11*, 3581.
(142) Shrock, E.; et al. Viral epitope profiling of COVID-19 patients reveals cross-reactivity and correlates of severity. *Science* **2020**, *370*, No. eabd4250.
(143) Roxhed, N.; et al. Multianalyte serology in home-sampled blood enables an unbiased assessment of the immune response against SARS-CoV-2. *Nat. Commun.* **2021**, *12*, 3695.
(144) Houzé, P.; et al. New Trend in Toxicological Screening Using Volumetric Absorptive Microsampling (VAMS) and High-Resolution Mass Spectrometry (HR/MS) Combination. *Molecules* **2023**, *28*, 3466.
(145) Lei, B. U. W.; Prow, T. W. A review of microsampling techniques and their social impact. *Biomed. Microdevices* **2019**, *21*, 81.
(146) Ribet, F.; et al. Microneedle Patch for Painless Intradermal Collection of Interstitial Fluid Enabling Multianalyte Measurement of Small Molecules, SARS-CoV-2 Antibodies, and Protein Profiling. *Adv. Healthcare Mater.* **2023**, *12*, No. e2202564.
(147) Shen, X.; et al. Multi-omics microsampling for the profiling of lifestyle-associated changes in health. *Nat. Biomed. Eng.* **2024**, *8*, 11.
(148) Fredolini, C.; et al. Proteome profiling of home-sampled dried blood spots reveals proteins of SARS-CoV-2 infections. *Commun. Med.* **2024**, *4*, 1−13.
(149) Motone, K.; Nivala, J. Not if but when nanopore protein sequencing meets single-cell proteomics. *Nat. Methods* **2023**, *20*, 336−338.
(150) Tognetti, M.; et al. Biomarker Candidates for Tumors Identified from Deep-Profiled Plasma Stem Predominantly from the Low Abundant Area. *J. Proteome Res.* **2022**, *21*, 1718−1735.
(151) Fulton, R. J.; McDade, R. L.; Smith, P. L.; Kienker, L. J.; Kettman, J. R. Advanced multiplexed analysis with the FlowMetrix system. *Clin. Chem.* **1997**, *43*, 1749−1756.
(152) Schwenk, J. M.; et al. Toward next generation plasma profiling via heat-induced epitope retrieval and array-based assays. *Mol. Cell. Proteomics MCP* **2010**, *9*, 2497−2507.
(153) Egertson, J. D.; et al. A theoretical framework for proteome-scale single-molecule protein identification using multi-affinity protein binding reagents. 2021; bioRxiv preprint at DOI: 10.1101/2021.10.11.463967.
(154) Rissin, D. M.; et al. Single-molecule enzyme-linked immunosorbent assay detects serum proteins at subfemtomolar concentrations. *Nat. Biotechnol.* **2010**, *28*, 595−599.
(155) Reed, B. D.; et al. Real-time dynamic single-molecule protein sequencing on an integrated semiconductor device. *Science* **2022**, *378*, 186−192.
(156) Blume, J. E.; et al. Rapid, deep and precise profiling of the plasma proteome with multi-nanoparticle protein corona. *Nat. Commun.* **2020**, *11*, 3662.
(157) Candia, J.; Daya, G. N.; Tanaka, T.; Ferrucci, L.; Walker, K. A. Assessment of variability in the plasma 7k SomaScan proteomics assay. *Sci. Rep.* **2022**, *12*, 17147.
(158) Anderson, N. L.; Anderson, N. G.; Haines, L. R.; Hardie, D. B.; Olafson, R. W.; Pearson, T. W. Mass Spectrometric Quantitation of Peptides and Proteins Using Stable Isotope Standards and Capture by Anti-Peptide Antibodies (SISCAPA). *J. Proteome Res.* **2004**, *3*, 235−244.
(159) Oh, H. S.-H.; et al. Organ aging signatures in the plasma proteome track health and disease. *Nature* **2023**, *624*, 164−172.
(160) Carrasco-Zanini, J.; Wheeler, E.; Uluvar, B.; Kerrison, N.; Koprulu, M.; Wareham, N. J.; Pietzner, M.; Langenberg, C. Mapping biological influences on the human plasma proteome beyond the genome. *Nat. Metab.* **2004**, 4714.
(161) Li, W.; et al. Role of exosomal proteins in cancer diagnosis. *Mol. Cancer* **2017**, *16*, 145.
(162) Wainberg, M.; et al. Multiomic blood correlates of genetic risk identify presymptomatic disease alterations. *Proc. Natl. Acad. Sci. U. S. A.* **2020**, *117*, 21813−21820.
(163) Bader, J. M. Proteome profiling in cerebrospinal fluid reveals novel biomarkers of Alzheimer's disease. *Mol. Syst. Biol.* **2020**, *16*, e9356.
(164) Demichev, V.; et al. A time-resolved proteomic and prognostic map of COVID-19. *Cell Syst.* **2021**, *12*, 780−794.e7.
(165) Bauer, W.; et al. Plasma Proteome Fingerprints Reveal Distinctiveness and Clinical Outcome of SARS-CoV-2 Infection. *Viruses* **2021**, *13*, 2456.
(166) Zhang, Y.-H.; et al. Lung proteomic biomarkers associated with chronic obstructive pulmonary disease. *Am. J. Physiol.-Lung Cell. Mol. Physiol.* **2021**, *321*, L1119−L1130.
(167) Álvez, M. B.; et al. Next generation pan-cancer blood proteome profiling using proximity extension assay. *Nat. Commun.* **2023**, *14*, 4308.
(168) Palma Medina, L. M.; et al. Targeted plasma proteomics reveals signatures discriminating COVID-19 from sepsis with pneumonia. *Respir. Res.* **2023**, *24*, 62.
(169) Luo, H.; et al. Associations of plasma proteomics with type 2 diabetes and related traits: results from the longitudinal KORA S4/F4/FF4 Study. *Diabetologia* **2023**, *66*, 1655−1668.
(170) Leuschner, G.; et al. Mass spectrometry-based autoimmune profiling reveals predictive autoantigens in idiopathic pulmonary fibrosis. *iScience* **2023**, *26*, No. 108345.
(171) Åkesson, J.; et al. Proteomics reveal biomarkers for diagnosis, disease activity and long-term disability outcomes in multiple sclerosis. *Nat. Commun.* **2023**, *14*, 6903.
(172) Arakawa, N.; et al. Stratifin as a novel diagnostic biomarker in serum for diffuse alveolar damage. *Nat. Commun.* **2022**, *13*, 5854.
(173) Yin, Y.; et al. Proteome profiling of early gestational plasma reveals novel biomarkers of congenital heart disease. *EMBO Mol. Med.* **2023**, *15*, No. e17745.
(174) Xiong, Y.; et al. Circulating proteomic panels for risk stratification of intracranial aneurysm and its rupture. *EMBO Mol. Med.* **2022**, *14*, No. e14713.
(175) Surinova, S.; et al. Prediction of colorectal cancer diagnosis based on circulating plasma proteins. *EMBO Mol. Med.* **2015**, *7*, 1166−1178.

https://doi.org/10.1021/acs.jproteome.4c00586
J. Proteome Res. XXXX, XXX, XXX−XXX

(176) Hoshino, A.; et al. Extracellular Vesicle and Particle Biomarkers Define Multiple Human Cancers. *Cell* **2020**, *182*, 1044−1061.e18.
(177) Kotol, D.; et al. Absolute Quantification of Pan-Cancer Plasma Proteomes Reveals Unique Signature in Multiple Myeloma. *Cancers* **2023**, *15*, 4764.
(178) Lightbody, E. D.; et al. High-Throughput Plasma Proteomics to Define the Precursor Multiple Myeloma Proteome and Identify Candidate High-Risk Disease Biomarkers of Progression. *Blood* **2023**, *142*, 334.
(179) Ramonfaur, D.; et al. High Throughput Plasma Proteomics and Risk of Heart Failure and Frailty in Late Life. *JAMA Cardiol.* **2024**, *9*, 649.

https://doi.org/10.1021/acs.jproteome.4c00586
J. Proteome Res. XXXX, XXX, XXX−XXX