



Towards Building a Quantitative Proteomics Toolbox in Precision Medicine: A Mini-Review

Alejandro Correa Rojo^{1,2*}, Dries Heylen^{1,2}, Jan Aerts¹, Olivier Thas^{1,3,4}, Jef Hooyberghs^{2,5}, Gökhan Ertaylan^{2*} and Dirk Valkenborg^{1*}

¹Data Science Institute, Interuniversity Institute for Biostatistics and Statistical Bioinformatics (I-BioStat), Hasselt University, Diepenbeek, Belgium, ²Flemish Institute for Technological Research (VITO), Mol, Belgium, ³Department of Applied Mathematics, Computer Science and Statistics, Faculty of Sciences, Ghent University, Ghent, Belgium, ⁴National Institute for Applied Statistics Research Australia (NIASRA), Wollongong, NSW, Australia, ⁵Theoretical Physics, Data Science Institute, Hasselt University, Diepenbeek, Belgium

OPEN ACCESS

Edited by:

Matteo Barberis,
University of Surrey, United Kingdom

Reviewed by:

Jussi Tapani Koivumäki,
Tampere University, Finland
John S. Clemmer,
University of Mississippi Medical
Center, United States
Sumio Ohtsuki,
Kumamoto University, Japan

*Correspondence:

Alejandro Correa Rojo
alejandro.correarojo@uhasselt.be
Gökhan Ertaylan
gokhan.ertaylan@vito.be
Dirk Valkenborg
dirk.valkenborg@uhasselt.be

Specialty section:

This article was submitted to
Computational Physiology and
Medicine,
a section of the journal
Frontiers in Physiology

Received: 10 June 2021

Accepted: 05 August 2021

Published: 26 August 2021

Citation:

Correa Rojo A, Heylen D, Aerts J,
Thas O, Hooyberghs J,
Ertaylan G and Valkenborg D (2021)
Towards Building a Quantitative
Proteomics Toolbox in Precision
Medicine: A Mini-Review.
Front. Physiol. 12:723510.
doi: 10.3389/fphys.2021.723510

Precision medicine as a framework for disease diagnosis, treatment, and prevention at the molecular level has entered clinical practice. From the start, genetics has been an indispensable tool to understand and stratify the biology of chronic and complex diseases in precision medicine. However, with the advances in biomedical and omics technologies, quantitative proteomics is emerging as a powerful technology complementing genetics. Quantitative proteomics provide insight about the dynamic behaviour of proteins as they represent intermediate phenotypes. They provide direct biological insights into physiological patterns, while genetics accounting for baseline characteristics. Additionally, it opens a wide range of applications in clinical diagnostics, treatment stratification, and drug discovery. In this mini-review, we discuss the current status of quantitative proteomics in precision medicine including the available technologies and common methods to analyze quantitative proteomics data. Furthermore, we highlight the current challenges to put quantitative proteomics into clinical settings and provide a perspective to integrate proteomics data with genomics data for future applications in precision medicine.

Keywords: precision medicine, quantitative proteomics, targeted techniques, bioinformatics, biomarker discovery, clinical diagnostics, protein quantitative trait loci

INTRODUCTION

Precision medicine aims to stratify patient populations so as to provide targeted and efficient treatments and reduce adverse treatment effects for human health (König et al., 2017). Furthermore, it brings opportunities for the healthcare industry by utilizing novel diagnostics platforms and specialized treatments that combine large-scale data with high-end computational analyses (Flores et al., 2013; Siwy et al., 2019).

The advances of biomedical and molecular technologies reduced per-individual cost of high-throughput technologies, such as next-generation sequencing and targeted proteomics. These advances bring omics sciences as a feasible approach to unravel molecular patterns of disease and wellbeing, and hence put precision medicine into clinical practice (Olivier et al., 2019; Morello et al., 2020). Genomics has been the most used approach given the high amount

of available genetic data and its association with traits and chronic diseases, such as cancer (Malone et al., 2020), type II diabetes mellitus (T2D; Scott et al., 2017), and cardiometabolic diseases (Dainis and Ashley, 2018). Still, most genetic studies provide associations between genes and risks for a disease, no direct mechanistic markers are found that explain the disease etiology, expediting the need to associate with other molecular layers and environmental factors (Tam et al., 2019). Despite great scientific and technological developments in recent years, many applications are still at the research-grade level requiring demonstration of clinical validation and usability (Liu et al., 2019).

Proteomics is the next likely candidate to be included in the precision medicine arsenal, for proteins represent intermediate phenotypes. In particular, proteins are products of gene expression and mediate biochemical activities of cells and tissues (Ding et al., 2019). Proteomics approaches could describe disease-related pathways; identify novel biomarkers for diagnostics; detect drug targets; and analyze physiological patterns on the transition for disease (Van Eyk and Snyder, 2018).

More specifically, quantitative proteomics has emerged as an important technique for precision medicine because it provides information about the physiological differences between biological samples based on the protein abundance levels. Thus, quantitative proteomics has relevant applications for the clinical and biomedical field including biomarker and drug discovery (Prasad et al., 2017). For the detection of human proteins, targeted approaches are often used which include targeted mass spectrometry (MS) techniques or affinity-reagent-based platforms. Targeted techniques aim to quantify the abundance of preselected proteins from an individual and thus correlate concentration values with patterns of disease.

Mass spectrometry is the most common technique in proteomics studies and has been widely used to measure proteins in the blood. Recent innovations in MS techniques have brought novel methods to measure human proteins, such as data-independent acquisition (DIA) methods and mass spectrometry imaging (MSI). DIA methods combine the reproducibility of single/parallel/multiple reaction monitoring with the high-throughput discovery aspect of shotgun proteomics while remaining comprehensive (Zhang et al., 2020). Conversely, MSI is transforming pathology allowing to identify precise and quantitative changes of proteins across individuals, disease states, tissues, and time (Ščupáková et al., 2020). Up to date, targeted MS-based blood proteomics have detected more than 17,000 proteins from coding genes in the human proteome (Kim et al., 2014; Adhikari et al., 2020). Yet, implementations for the human blood proteome in clinical settings are limited because targeted MS techniques require multiple sample preparation steps including removal of high-abundance proteins, trypsin digestion, and liquid chromatography (Maes et al., 2015).

Affinity-based methods have been considered as an alternative approach to MS. These are often based on antibodies to target specific proteins in a biological sample and they are considered the gold standard for clinical diagnostics. Classical techniques, such as ELISA, use polyclonal or monoclonal antibodies to capture protein targets (Brennan et al., 2010). However, due to their cross-reactivity, they have poor specificity, and sensitivity

for low-abundant proteins in human samples, and they are consequently not suitable for high content, large-scale analyses, or high coverage of human proteins (Ellington et al., 2010).

With the advances of multiplexing technologies, immunoassays techniques have been improved for simultaneously measuring multiple proteins with a wide range of concentrations in multiple samples. Compared to targeted MS, multiplexed technologies are high throughput, have high sensitivity for low abundant proteins, and target specific proteins of clinical relevance (Smith and Gerszten, 2017). Therefore, in this mini-review, we will briefly discuss current innovations and applications of quantitative proteomics, based on high-throughput multiplexed technologies, in precision medicine and their current status in the clinic (Figure 1).

MULTIPLEXED AFFINITY-REAGENT-BASED METHODS

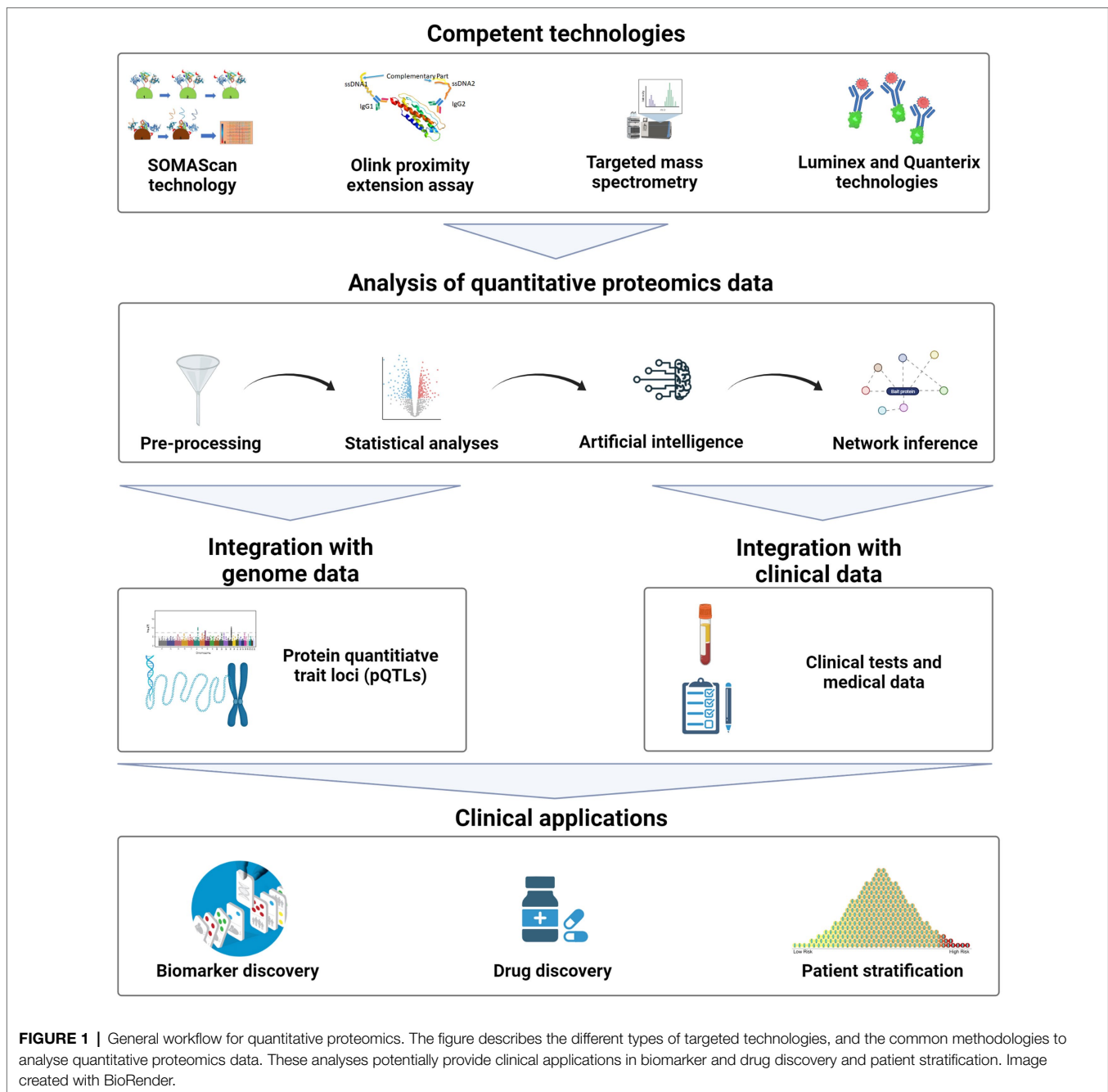
Multiplexed immunoassay technologies include improved binding reagents to increase affinity and specificity, using multiplexed ELISA arrays (Luminex and Quanterix), antibody labeled nucleotides (Olink), or aptamers (SOMAScan).

Luminex and Quanterix (SIMOA) technologies are based on suspension bed arrays in which captured antibodies are attached to different fluorescent-dyed microparticles. Each colored microparticle represents one assay for a given protein target. Proteins are then measured by flow cytometry analysis (Tait et al., 2009; Rissin et al., 2010). These techniques can quantify up to 50 proteins and process up to 384 samples in batches (Wilson et al., 2016).

Conversely, Olink technology (Olink Proteomics) uses antibodies that are labeled with nucleotides and detect proteins in a sample by proximity extension assay (PEA). Antibodies that are linked with complementary oligonucleotides which upon binding the target protein, the oligonucleotides are hybridized and then extended using a DNA polymerase. The initial concentration of the protein target is measured by the concentration of the generated DNA amplicon, using quantitative PCR (Assarsson et al., 2014). Nowadays, the platform can detect up to 1,162 clinically relevant proteins distributed across 15 protein panels related to cardiometabolic disorders, cell regulation, cardiovascular diseases, immune system, oncology, inflammation, metabolism, and neurology. Additionally, each panel allows multiplexing for 90 samples per batch.

SOMAScan technology (SOMALogic) uses aptamers to achieve high sensitivity and high multiplexing. Aptamers are short oligonucleotides developed by a pool of random sequence oligomers that binds to a target protein. Captured proteins by aptamers are then measured using a DNA microarray (Gold et al., 2010). The current version of this platform can measure more than 7,000 proteins and processes 90 samples per batch.

Compared to MS techniques, multiplexed affinity-reagent-based methods achieve high coverage, high sensitivity for several low abundances, high specificity for target proteins, and good reproducibility (low intra-assay coefficient variation; Smith and Gerszten, 2017; Petrerá et al., 2021). However, they have several



limitations which include as: no detection of proteins that are not targeted by the assay (comprehensiveness); binding affinity differences across proteins or non-specific binding for variant proteins (quantitative accuracy); and no distinction between posttranslational modified proteins and isoforms (specificity; Yeh et al., 2017; Raffield et al., 2020). For SOMAScan and Olink, Pietzner et al. (2021) showed that factors of technical variability can be introduced by target proteins with transmembrane domain, glycosylation effects, or protein-altering variants (Pietzner et al., 2021). Still, implementations of multiplexed platforms into clinical settings are relatively new, given that more research and verification are still needed to

validate these as clinical-grade technologies (Williams et al., 2019). For more information about the recent technical validation of these platforms, we encourage the readers to review the work of Petrer et al. (2021) and Pietzner et al. (2021).

Nevertheless, multiplexed affinity-based methods are now been used for large-population analyses to link proteomics data with genomic data. Affinity-based assays provide a direct link between protein levels and genetic variants which can unravel causes of complex traits and detect biological effects on the protein layer. We provide a summarized table with the current large-population cohorts using these techniques (Table 1).

TABLE 1 | Summary of large-scale population cohorts with quantitative proteomics data.

Study population	Research topic	Number of samples	Number of proteins measured	Type of sample	Number of associated proteins with genetic effects	Platform	Reference
KORA and QMDiab study	Association between genetic risk and plasma proteins for diseases	1,335	1,124	Plasma	284	SOMAScan	Suhre et al., 2017
INTERVAL study	Association study between genetic variants and the human plasma proteome	3,301	3,622	Plasma	1,478	SOMAScan Olink Proteomics	Sun et al., 2018
AGES Reykjavik study	Epidemiologic study to examine risk factors and gene–environment interactions for disease and disability in old age	5,457	4,137	Serum	2,148	SOMAScan	Emilsson et al., 2018
LifeLines Dutch population cohort	Multi-generational cohort study to study the etiology of several diseases	1,264	92	Plasma/ Serum	74	Olink Proteomics	LifeLines cohort study et al., 2018
SCALLOP consortium	Collaborative framework across multiple studies to map pQTLs and analyze biomarker proteins on the Olink proteomics platform	30,931	90	Plasma/ Serum	85 ¹	Olink Proteomics	Folkersen et al., 2020
Latin Americans ORIGIN study	Genetic study on the human serum proteome for novel biomarkers in cardiovascular diseases	2,216	237	Serum	23	Luminex immunoassay	Sjaarda et al., 2020

¹The number of detected proteins showed in the table represents the first published results of the SCALLOP consortium. More population cohorts are being added to the consortium which currently comprises data from 71,232 samples.

ANALYSIS OF QUANTITATIVE PROTEOMICS DATA

The analysis of quantitative proteomics data is quite challenging. Depending on the targeted technology used, experimental design, and the type of research question being addressed, specific computational workflows are needed. Bioinformatics has provided a wide range of methods, not only to analyze large-scale proteomics data but also to integrate it with other types of omics data for clinical research. However, standardized workflows are needed to successfully put quantitative proteomics analyses into clinical practice (Martens, 2013). In this section, we review the common and promising methods for analyzing proteomics data based on large-scale studies.

Data Pre-processing

In omics data analysis, bias refers to systematic features of the data that can be attributed to experimental and/or technical factors that are related to sample preparation, the platform runs, data acquisition, etc. Normalization is the process that aims to correct such biases (Välikangas et al., 2016). In comparison with targeted MS techniques, normalization in multiplex affinity-reagent-based methods is relatively straightforward. The main assumption on these techniques is that protein levels are measured based on targeted antigen/antibody affinity-binding. This implies that abundance levels are not influenced by factors that cause protein isoforms, such as, posttranslational modifications or spliced variants. However, as mentioned before, recent studies have shown biological variations that interfere with the analysis of the data which require further research on pre-processing methods. Nevertheless, we discuss the current approaches used for quantitative proteomics data.

Before normalization, traditional quantitative proteomics data must be transformed to adjust for the effect of protein levels and detect changes in abundances between samples (Quackenbush, 2002). Several methods exist but the most frequently used is the \log_2 transformation because it allows easy interpretation of fold change in protein levels (Karpievitch et al., 2012). After transformation, normalization is applied. The most common methods derived from MS techniques or microdata array methodologies include global and quantile normalization (Bolstad et al., 2003; Chawade et al., 2014), regression models (Callister et al., 2007), and constrained optimization, such as CONSTANd (Maes et al., 2016). However, for Olink and SOMAScan, the pre-processing starts from normalization as the manufacturers provide their normalization guidelines. For Olink, data are normalized based on normalized protein expression values (NPX) (Sun et al., 2018; Zhong et al., 2021) while for SOMAScan, data are normalized by estimating relative fluorescence intensities (RFUs; Candia et al., 2017).

Batch effects are also an important consideration in data pre-processing. Although normalization methods aim to correct for these effects simultaneously, some sources of variations are resistant to these approaches. For large proteomics datasets, empirical Bayes methods, such as ComBat (Johnson et al., 2007; Leek et al., 2012), have been used to adjust for known batch effects (Kim et al., 2018; Kalla et al., 2021).

Despite the availability of multiple pre-processing methods for quantitative proteomics data, the main limitation is the lack of methodologies to compare protein levels between multiple cohorts. The application of the previously mentioned methods is not yet fully studied and transparently communicated. Validation of these methods for affinity-based

techniques is necessary to compare data from multiple targeted platforms and obtain reproducible results (Rausch et al., 2016).

Statistical and Enrichment Analyses

Traditional statistical analyses compare protein levels between study groups or conditions and detect which proteins are significantly differentially expressed. This is commonly done by performing two-sample *t*-tests between protein abundances or an ANOVA when two or more conditions are to be compared (Kammers et al., 2015). For more robust and accurate results, Linear Models for Microarray Data (LIMMA) are used (Ritchie et al., 2015).

For large-scale proteomics analyses, multiple hypotheses are being tested which is necessary to control for false positives. Statistical estimates, such as false discovery rate and the Benjamini-Hochberg procedure (BH), are used to obtain true biological results (Aggarwal and Yadav, 2016; Korthauer et al., 2019).

In addition to the previously mentioned methods, Olink Proteomics offers an open-source toolbox, OlinkAnalyze, to pre-process and do quick analyses for Olink's data.¹ Conversely, SOMALogic also provides a platform for the pre-processing and analysis of aptamer-based proteomics data.²

Results from statistical analyses do not yet provide the biological context of differentially expressed proteins. To understand the functional features and effects of the detected proteins, an enrichment analysis must be performed. This helps to generate hypotheses on the systemic response of the proteome, revealing and understanding the biological processes that underlie the quantitative profiles of the proteins. Methods include simple classification of proteins using large public databases, such as UniProt (The UniProt Consortium et al., 2021) and Ensembl (Howe et al., 2021), and Gene Ontology (GO) analyses from resources, such as AmiGO database (Carbon et al., 2009); EggNOG (Jensen et al., 2007); and MetaCore™.

Artificial Intelligence-Based Methods

Artificial intelligence-based methods can extend traditional statistical analyses by extracting informative features and building models that can predict or describe relevant outcomes. Using supervised and unsupervised techniques, a variety of models include Random Forest, support vector machines (SVMs), Artificial Neural Network, regression models, and K-means clustering (Chen et al., 2020). In quantitative proteomics, based on multiplexed affinity-reagent-based methods, these techniques have been used to predict disease signatures or clinical outcomes. Suvarna et al. (2021) identified protein classifiers of patients with non-severe and severe COVID-19, by using SVMs models (Suvarna et al., 2021). Hewitson et al. (2021) used Random Forest and logistic regression models to classify proteins in blood as potential biomarkers in autism spectrum disorder (Hewitson et al., 2021).

¹<https://github.com/Olink-Proteomics/OlinkRPackage>

²<https://github.com/SomaLogic/SomaLogic-Data>

Network Inference

Mapping interactions and associations between different proteins allow presenting proteomics data as networks. These interactions reflect molecular entities as building blocks of any type of biological process, especially signaling, regulation, and biochemical interactions. Two distinct strategies of network inference are possible. Validated pathways and mechanisms can be consulted in resources, such as KEGG (Kanehisa et al., 2016), ENCODE (The ENCODE Project Consortium, 2012), PathVisio (Kutmon et al., 2015), MetaCore™, WikiPathways (Slenter et al., 2018), Reactome (Jassal et al., 2019), BioGrid (Stark, 2006), STRING (Jensen et al., 2009), and iPathwayGuide™. Such knowledge-based approach can guide integrative analyses by making use of established information from validated experiments, databases, and scientific literature.

In a more data-driven approach, statistical or machine learning methods can be used for inferring relationships, correlating between proteins and/or other molecules, and exploring novel interactions. Common methods include weighted gene co-expression network analysis, Gaussian graphical models, Bayesian networks, and Markov Chain Monte Carlo (MCMC; Mohammadi and Wit, 2015; Hawe et al., 2019).

INTEGRATION WITH GENOMIC DATA

Genomics have always been the key technology in personalized medicine. Genome-wide association studies (GWAS) have been used to test millions of genetic variants across many individuals to identify genotype–phenotype associations. Overall, more than 50,000 associations have been reported between genetic variants, common diseases, and traits (Loos, 2020). However, GWAS has not been able to bridge the gap between genotype and phenotype because most of the identified associations only explain a small fraction of heritability and do not provide causality between genetic variants and traits.

Quantitative proteomics can extend GWAS toward proteome-wide association studies (PWAS) by studying protein quantitative trait loci (pQTLs). pQTLs refer to associations between genetic variants and protein abundance levels which can be *cis*-pQTLs or *trans*-pQTLs (Suhre et al., 2021). *Cis*-pQTLs specify variants that are likely to have a direct effect on the observed protein levels at that locus, whereas *trans*-pQTLs specify a variant distant to the protein-coding gene or on another chromosome that could indicate an indirect link (Molendijk and Parker, 2021).

In the context of precision medicine, several studies have successfully described phenotypic features of complex diseases using PWAS. Wingo et al. (2021) integrated 376 human brain proteomes with GWAS data from 455,528 individuals in which 13 coding genes were found causal for protein levels as well to be correlated with Alzheimer's disease, neuroticism, and Parkinson disease (Wingo et al., 2021). Zaghlool et al. (2021) studied the association between 1,000 plasma proteins and body mass index over 4,600 participants where 21 proteins in pathways of adiposity were found to be causal drivers in obesity-associated pathologies (Zaghlool et al., 2021).

CLINICAL APPLICATIONS

The applications for quantitative proteomics in precision medicine are numerous. Proteomics promises to contribute to the stratification of treatment options for patients. It can provide robust support for biomarker discovery and drug development. Additionally, it can be integrated with genetic data to support genetic risk scores for complex diseases. Before these potential applications of quantitative proteomics can be realized, an important consideration is that proteomics data may reveal personal data. Hence, ethical, privacy, and data sharing frameworks are needed to allow secured research in precision medicine (Boonen et al., 2019). Below, we highlight three promising applications of quantitative proteomics in the clinic.

Diagnostics, Biomarker Discovery, and Surrogate End-Points

In general, most proteomics studies in the clinic are aimed at the identification of biomarkers that are specific for the diagnosis of disease or associated with disease severity. Recent studies have identified potential biomarkers for different types of disease. Franzén et al. (2021) identified 33 protein biomarkers of non-small-cell lung cancer related to different stages of disease severity (Franzén et al., 2021). Sonnenschein et al. (2021) identified c-KIT as a novel biomarker from serum proteins to distinguish between patients with hypertrophic cardiomyopathy and healthy subjects (Sonnenschein et al., 2021).

Pharmacoproteomics

Integration of genomic data in large-scale proteomics studies is now providing novel methodologies for drug target identification. With the ongoing research on pQTLs, recent GWAS and PWAS have identified potential drug targets for several diseases. From one UK Biobank study, Bretherick et al. (2020) detected 38 proteins with pQTL effects in inflammatory bowel disease, coronary artery disease, and schizophrenia. From these proteins, 1,319 compounds were associated as potential therapeutic agents (Bretherick et al., 2020).

REFERENCES

- Adhikari, S., Nice, E. C., Deutsch, E. W., Lane, L., Omenn, G. S., Pennington, S. R., et al. (2020). A high-stringency blueprint of the human proteome. *Nat. Commun.* 11:5301. doi: 10.1038/s41467-020-19045-9
- Aggarwal, S., and Yadav, A. K. (2016). False discovery rate estimation in proteomics. *Methods Mol. Biol.* 1362, 119–128. doi: 10.1007/978-1-4939-3106-4_7
- Assarsson, E., Lundberg, M., Holmquist, G., Björkstén, J., Bucht Thorsen, S., Ekman, D., et al. (2014). Homogenous 96-Plex PEA immunoassay exhibiting high sensitivity, specificity, and excellent scalability. *PLoS One* 9:e95192. doi: 10.1371/journal.pone.0095192
- Benson, M. D., Yang, Q., Ngo, D., Zhu, Y., Shen, D., Farrell, L. A., et al. (2018). Genetic architecture of the cardiovascular risk proteome. *Circulation* 137, 1158–1172. doi: 10.1161/CIRCULATIONAHA.117.029536
- Bolstad, B. M., Irizarry, R. A., Astrand, M., and Speed, T. P. (2003). A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* 19, 185–193. doi: 10.1093/bioinformatics/19.2.185
- Boonen, K., Hens, K., Menschaert, G., Baggerman, G., Valkenburg, D., and Ertaylan, G. (2019). Beyond genes: re-Identifiability of proteomic data and its implications for personalized medicine. *Gen. Dent.* 10:682. doi: 10.3390/genes10090682
- Brennan, D. J., O'Connor, D. P., Rexhepaj, E., Ponten, F., and Gallagher, W. M. (2010). Antibody-based proteomics: fast-tracking molecular diagnostics in oncology. *Nat. Rev. Cancer* 10, 605–617. doi: 10.1038/nrc2902
- Bretherick, A. D., Canela-Xandri, O., Joshi, P. K., Clark, D. W., Rawlik, K., Boutin, T. S., et al. (2020). Linking protein to phenotype with Mendelian randomization detects 38 proteins with causal roles in human diseases and traits. *PLoS Genet.* 16:e1008785. doi: 10.1371/journal.pgen.1008785
- Callister, S. J., Barry, R. C., Adkins, J. N., Johnson, E. T., Webb-Robertson, B.-J. M., Smith, R. D., et al. (2007). Normalization approaches for removing systematic biases associated with mass spectrometry and label-free proteomics. *J. Proteome Res.* 5, 277–286. doi: 10.1021/pr050300l
- Candia, J., Cheung, F., Kotliarov, Y., Fantoni, G., Sellers, B., Griesman, T., et al. (2017). Assessment of variability in the SOMAscan assay. *Sci. Rep.* 7:14248. doi: 10.1038/s41598-017-14755-5
- Carbon, S., Ireland, A., Mungall, C. J., Shu, S., Marshall, B., Lewis, S., et al. (2009). AmiGO: online access to ontology and annotation data. *Bioinformatics* 25, 288–289. doi: 10.1093/bioinformatics/btn615

Polygenic Risk Scores

Polygenic risk scores (PRSs) are a novel approach to integrate individual genetic data into clinical settings. These scores aggregate the effect of multiple risk variants to assess the individual genetic predisposition for a given disease (Lewis and Vassos, 2020). Proteomics analyses can be embedded in PRSs, not only for novel biomarkers but also to assess the causes and prognosis of disease. Few studies for coronary artery disease and T2D have successfully integrated PRSs with protein levels which have provided novel associations between gene and protein levels as well as individual risk profiles for disease progression (Benson et al., 2018; Gudmundsdottir et al., 2020).

CONCLUSION

Quantitative proteomics is emerging as a powerful technology for precision medicine. For decades, MS has been the standard for quantitative proteomics for researchers, but new alternatives in affinity-reagent-based assays allow for high-throughput screening of proteins. Recent innovations provide tools for clinicians to medical applications, including in diagnostics, stratification, and treatment of diseases. However, substantial work is required for the validation of technologies, standardization of data analyses, and integration of proteomics with other molecular and phenotypic level data. Despite these challenges, recent progress is promising for the emerging quantitative proteomics toolbox to be used in clinical settings.

AUTHOR CONTRIBUTIONS

AR, DH, GE, and DV have equally contributed to the conceiving of the manuscript idea. AR and DH have drafted the manuscript with support from GE and DV. JA, JH, and OT have provided critical comments on the draft manuscript. All authors read and approved the final manuscript.

- Chawade, A., Alexandersson, E., and Levander, F. (2014). Normalyzer: A tool for rapid evaluation of normalization methods for omics data sets. *J. Proteome Res.* 13, 3114–3120. doi: 10.1021/pr401264n
- Chen, C., Hou, J., Tanner, J. J., and Cheng, J. (2020). Bioinformatics methods for mass spectrometry-based proteomics data analysis. *Int. J. Mol. Sci.* 21:2873. doi: 10.3390/ijms21082873
- Dainis, A. M., and Ashley, E. A. (2018). Cardiovascular precision medicine in the genomics era. *JACC Basic Transl. Sci.* 3, 313–326. doi: 10.1016/j.jacbs.2018.01.003
- Ding, C., Qin, Z., Li, Y., Shi, W., Li, L., Zhan, D., et al. (2019). Proteomics and precision medicine. *Small Methods* 3:1900075. doi: 10.1002/smt.201900075
- Ellington, A. A., Kullo, I. J., Bailey, K. R., and Klee, G. G. (2010). Antibody-based protein multiplex platforms: technical and operational challenges. *Clin. Chem.* 56, 186–193. doi: 10.1373/clinchem.2009.127514
- Emilsson, V., Ilkov, M., Lamb, J. R., Finkel, N., Gudmundsson, E. F., Pitts, R., et al. (2018). Co-regulatory networks of human serum proteins link genetics to disease. *Science* 361, 769–773. doi: 10.1126/science.aaq1327
- Flores, M., Glusman, G., Brogaard, K., Price, N. D., and Hood, L. (2013). P4 medicine: how systems medicine will transform the healthcare sector and society. *Per. Med.* 10, 565–576. doi: 10.2217/pme.13.57
- Folkersen, L., Gustafsson, S., Wang, Q., Hansen, D. H., Hedman, Å. K., Schork, A., et al. (2020). Genomic and drug target evaluation of 90 cardiovascular proteins in 30,931 individuals. *Nat. Metab.* 2, 1135–1148. doi: 10.1038/s42255-020-00287-2
- Franzén, B., Viktorsson, K., Kamali, C., Darai-Ramqvist, E., Grozman, V., Arapi, V., et al. (2021). Multiplex immune protein profiling of fine-needle aspirates from patients with non-small-cell lung cancer reveals signatures associated with PD-L1 expression and tumor stage. *Mol. Oncol.* 12952. doi: 10.1002/1878-0261.12952, [Epub ahead of print]
- Gold, L., Ayers, D., Bertino, J., Bock, C., Bock, A., Brody, E. N., et al. (2010). Aptamer-based multiplexed proteomic technology for biomarker discovery. *PLoS One* 5:e15004. doi: 10.1371/journal.pone.0015004
- Gudmundsdottir, V., Zaghlood, S. B., Emilsson, V., Aspelund, T., Ilkov, M., Gudmundsson, E. F., et al. (2020). Circulating protein signatures and causal candidates for type 2 diabetes. *Diabetes* 69, 1843–1853. doi: 10.2337/db19-1070
- Hawe, J. S., Theis, F. J., and Heinig, M. (2019). Inferring interaction networks from multi-omics data. *Front. Genet.* 10:535. doi: 10.3389/fgene.2019.00535
- Hewitson, L., Mathews, J. A., Devlin, M., Schutte, C., Lee, J., and German, D. C. (2021). Blood biomarker discovery for autism spectrum disorder: A proteomic analysis. *PLoS One* 16:e0246581. doi: 10.1371/journal.pone.0246581
- Howe, K. L., Achuthan, P., Allen, J., Allen, J., Alvarez-Jarreta, J., Amode, M. R., et al. (2021). Ensembl 2021. *Nucleic Acids Res.* 49, D884–D891. doi: 10.1093/nar/gkaa942
- Jassal, B., Matthews, L., Viteri, G., Gong, C., Lorente, P., Fabregat, A., et al. (2019). The reactome pathway knowledgebase. *Nucleic Acids Res.* 48:gzk1031. doi: 10.1093/nar/gzk1031
- Jensen, L. J., Julien, P., Kuhn, M., von Mering, C., Muller, J., Doerks, T., et al. (2007). eggNOG: automated construction and annotation of orthologous groups of genes. *Nucleic Acids Res.* 36, D250–D254. doi: 10.1093/nar/gkm796
- Jensen, L. J., Kuhn, M., Stark, M., Chaffron, S., Creevey, C., Muller, J., et al. (2009). STRING 8—A global view on proteins and their functional interactions in 630 organisms. *Nucleic Acids Res.* 37, D412–D416. doi: 10.1093/nar/gkn760
- Johnson, W. E., Li, C., and Rabinovic, A. (2007). Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* 8, 118–127. doi: 10.1093/biostatistics/kxj037
- Kalla, R., Adams, A. T., Bergemalm, D., Vatn, S., Kennedy, N. A., Ricanek, P., et al. (2021). Serum proteomic profiling at diagnosis predicts clinical course, and need for intensification of treatment in inflammatory bowel disease. *J. Crohn's Colitis* 15, 699–708. doi: 10.1093/ecco-jcc/jjaa230
- Kammers, K., Cole, R. N., Tiengwe, C., and Ruczinski, I. (2015). Detecting significant changes in protein abundance. *EuPA Open Proteom.* 7, 11–19. doi: 10.1016/j.euprot.2015.02.002
- Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., and Tanabe, M. (2016). KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.* 44, D457–D462. doi: 10.1093/nar/gkv1070
- Karpievitch, Y. V., Dabney, A. R., and Smith, R. D. (2012). Normalization and missing value imputation for label-free LC-MS analysis. *BMC Bioinf.* 13:S5. doi: 10.1186/1471-2105-13-S16-S5
- Kim, M.-S., Pinto, S. M., Getnet, D., Nirujogi, R. S., Manda, S. S., Chaerkady, R., et al. (2014). A draft map of the human proteome. *Nature* 509, 575–581. doi: 10.1038/nature13302
- Kim, C. H., Tworoger, S. S., Stampfer, M. J., Dillon, S. T., Gu, X., et al. (2018). Stability and reproducibility of proteomic profiles measured with an aptamer-based platform. *Sci. Rep.* 8:8382. doi: 10.1038/s41598-018-26640-w
- König, I. R., Fuchs, O., Hansen, G., von Mutius, E., and Kopp, M. V. (2017). What is precision medicine? *Eur. Respir. J.* 50:1700391. doi: 10.1183/13993003.00391-2017
- Korthauer, K., Kimes, P. K., Duvallet, C., Reyes, A., Subramanian, A., Teng, M., et al. (2019). A practical guide to methods controlling false discoveries in computational biology. *Genome Biol.* 20:118. doi: 10.1186/s13059-019-1716-1
- Kutmon, M., van Iersel, M. P., Bohler, A., Kelder, T., Nunes, N., Pico, A. R., et al. (2015). PathVisio 3: An extendable pathway analysis toolbox. *PLoS Comput. Biol.* 11:e1004085. doi: 10.1371/journal.pcbi.1004085
- Leek, J. T., Johnson, W. E., Parker, H. S., Jaffe, A. E., and Storey, J. D. (2012). The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* 28, 882–883. doi: 10.1093/bioinformatics/bts034
- Lewis, C. M., and Vassos, E. (2020). Polygenic risk scores: from research tools to clinical instruments. *Genome Med.* 12:44. doi: 10.1186/s13073-020-00742-5
- LifeLines cohort study BIOS consortium Zhernakova, D. V., Le, T. H., Kurilshikov, A., Atanasovska, B., et al. (2018). Individual variations in cardiovascular-disease-related protein levels are driven by genetics and gut microbiome. *Nat. Genet.* 50, 1524–1532. doi: 10.1038/s41588-018-0224-7
- Liu, X., Luo, X., Jiang, C., and Zhao, H. (2019). Difficulties and challenges in the development of precision medicine. *Clin. Genet.* 95, 569–574. doi: 10.1111/cge.13511
- Loos, R. J. F. (2020). 15 years of genome-wide association studies and no signs of slowing down. *Nat. Commun.* 11:5900. doi: 10.1038/s41467-020-19653-5
- Maes, E., Hadiwikarta, W. W., Mertens, I., Baggerman, G., Hooyberghs, J., and Valkenborg, D. (2016). CONSTAND: A normalization method for isobaric labeled spectra by constrained optimization. *Mol. Cell. Proteomics* 15, 2779–2790. doi: 10.1074/mcp.M115.056911
- Maes, E., Mertens, I., Valkenborg, D., Pauwels, P., Rolfo, C., and Baggerman, G. (2015). Proteomics in cancer research: are we ready for clinical practice? *Crit. Rev. Oncol. Hematol.* 96, 437–448. doi: 10.1016/j.critrevonc.2015.07.006
- Malone, E. R., Oliva, M., Sabatini, P. J. B., Stockley, T. L., and Siu, L. L. (2020). Molecular profiling for precision cancer therapies. *Genet. Med.* 12:8. doi: 10.1186/s13073-019-0703-1
- Martens, L. (2013). Bringing proteomics into the clinic: The need for the field to finally take itself seriously. *Proteomics Clin. Appl.* 7, 388–391. doi: 10.1002/prca.201300020
- Mohammadi, A., and Wit, E. C. (2015). Bayesian structure learning in sparse Gaussian graphical models. *Bayesian Anal.* 10, 109–138. doi: 10.1214/14-BA889
- Molendijk, J., and Parker, B. L. (2021). Proteome-wide systems genetics to identify functional regulators of complex traits. *Cell Sys.* 12, 5–22. doi: 10.1016/j.cels.2020.10.005
- Morello, G., Salomone, S., D'Agata, V., Conforti, F. L., and Cavallaro, S. (2020). From multi-omics approaches to precision medicine in amyotrophic lateral sclerosis. *Front. Neurosci.* 14:21, 33192262. doi: 10.3389/fnins.2020.577755
- Olivier, M., Asmis, R., Hawkins, G. A., Howard, T. D., and Cox, L. A. (2019). The need for multi-omics biomarker signatures in precision medicine. *Int. J. Mol. Sci.* 13:4781. doi: 10.3390/ijms20194781
- Petrera, A., von Toerne, C., Behler, J., Huth, C., Thorand, B., Hilgendorff, A., et al. (2021). Multiplatform approach for plasma proteomics: complementarity of Olink proximity extension assay technology to mass spectrometry-based protein profiling. *J. Proteome Res.* 20, 751–762. doi: 10.1021/acs.jproteome.0c00641
- Pietzner, M., Wheeler, E., Carrasco-Zanini, J., Kerrison, N. D., Oerton, E., Koprulu, M., et al. (2021). Cross-platform proteomics to advance genetic prioritisation strategies. *BioRxiv [Preprint]*. doi:10.1101/2021.03.18.435919
- Prasad, B., Vrana, M., Mehrotra, A., Johnson, K., and Bhatt, D. K. (2017). The promises of quantitative proteomics in precision medicine. *J. Pharm. Sci.* 106, 738–744. doi: 10.1016/j.xphs.2016.11.017
- Quackenbush, J. (2002). Microarray data normalization and transformation. *Nat. Genet.* 32, 496–501. doi: 10.1038/ng1032

- Raffield, L. M., Dang, H., Pratte, K. A., Jacobson, S., Gillenwater, L. A., Ampleford, E., et al. (2020). Comparison of proteomic assessment methods in multiple cohort studies. *Proteomics* 20:1900278. doi: 10.1002/pmic.201900278
- Rausch, T. K., Schillert, A., Ziegler, A., Lüking, A., Zucht, H.-D., and Schulz-Knappe, P. (2016). Comparison of pre-processing methods for multiplex bead-based immunoassays. *BMC Genomics* 17:601. doi: 10.1186/s12864-016-2888-7
- Rissin, D. M., Kan, C. W., Campbell, T. G., Howes, S. C., Fournier, D. R., Song, L., et al. (2010). Single-molecule enzyme-linked immunosorbent assay detects serum proteins at subfemtomolar concentrations. *Nat. Biotechnol.* 28, 595–599. doi: 10.1038/nbt.1641
- Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., et al. (2015). Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 43:e47. doi: 10.1093/nar/gkv007
- Scott, R. A., Scott, L. J., Mägi, R., Marullo, L., Gaulton, K. J., Kaakinen, M., et al. (2017). An expanded genome-wide association study of type 2 diabetes in Europeans. *Diabetes* 66, 2888–2902. doi: 10.2337/db16-1253
- Ščupáková, K., Balluff, B., Tressler, C., Adelaja, T., Heeren, R. M. A., Glunde, K., et al. (2020). Cellular resolution in clinical MALDI mass spectrometry imaging: the latest advancements and current challenges. *Clin. Chem. Lab. Med.* 58, 914–929. doi: 10.1515/cclm-2019-0858
- Sivy, J., Mischak, H., and Zürbig, P. (2019). Proteomics and personalized medicine: A focus on kidney disease. *Expert Rev. Proteomics* 16, 773–782. doi: 10.1080/14789450.2019.1659138
- Sjaarda, J., Gerstein, H. C., Kutalik, Z., Mohammadi-Shemirani, P., Pigeire, M., et al. (2020). Influence of genetic ancestry on human serum proteome. *Am. J. Hum. Genet.* 106, 303–314. doi: 10.1016/j.ajhg.2020.01.016
- Slenter, D. N., Kutmon, M., Hanspers, K., Riutta, A., Windsor, J., Nunes, N., et al. (2018). WikiPathways: A multifaceted pathway database bridging metabolomics to other omics research. *Nucleic Acids Res.* 46, D661–D667. doi: 10.1093/nar/gkx1064
- Smith, J. G., and Gerszten, R. E. (2017). Emerging affinity-based proteomic Technologies for Large-Scale Plasma Profiling in cardiovascular disease. *Circulation* 135, 1651–1664. doi: 10.1161/CIRCULATIONAHA.116.025446
- Sonnenschein, K., Fiedler, J., de Gonzalo-Calvo, D., Xiao, K., Pfanne, A., Just, A., et al. (2021). Blood-based protein profiling identifies serum protein c-KIT as a novel biomarker for hypertrophic cardiomyopathy. *Sci. Rep.* 11:1755. doi: 10.1038/s41598-020-80868-z
- Stark, C. (2006). BioGRID: A general repository for interaction datasets. *Nucleic Acids Res.* 34, D535–D539. doi: 10.1093/nar/gkj109
- Suhre, K., Arnold, M., Bhagwat, A. M., Cotton, R. J., Engelke, R., Raffler, J., et al. (2017). Connecting genetic risk to disease end points through the human blood plasma proteome. *Nat. Commun.* 8:14357. doi: 10.1038/ncomms14357
- Suhre, K., McCarthy, M. I., and Schwenk, J. M. (2021). Genetics meets proteomics: perspectives for large population-based studies. *Nat. Rev. Genet.* 22, 19–37. doi: 10.1038/s41576-020-0268-2
- Sun, B. B., Maranville, J. C., Peters, J. E., Stacey, D., Staley, J. R., Blackshaw, J., et al. (2018). Genomic atlas of the human plasma proteome. *Nature* 558, 73–79. doi: 10.1038/s41586-018-0175-2
- Suvarna, K., Biswas, D., Pai, M. G. J., Acharjee, A., Bankar, R., Palanivel, V., et al. (2021). Proteomics and machine learning approaches reveal a set of prognostic markers for COVID-19 severity With drug repurposing potential. *Front. Physiol.* 12:652799. doi: 10.3389/fphys.2021.652799
- Tait, B. D., Hudson, F., Cantwell, L., Brewin, G., Holdsworth, R., Bennett, G., et al. (2009). Review article: Luminex technology for HLA antibody detection in organ transplantation. *Nephrology* 14, 247–254. doi: 10.1111/j.1440-1797.2008.01074.x
- Tam, V., Patel, N., Turcotte, M., Bossé, Y., Paré, G., and Meyre, D. (2019). Benefits and limitations of genome-wide association studies. *Nat. Rev. Genet.* 20, 467–484. doi: 10.1038/s41576-019-0127-1
- The ENCODE Project Consortium (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature* 489, 57–74. doi: 10.1038/nature11247
- The UniProt Consortium/Bateman, A., Martin, M.-J., Orchard, S., Magrane, M., Agivetova, R., et al. (2021). UniProt: The universal protein knowledgebase in 2021. *Nucleic Acids Res.* 49, D480–D489. doi: 10.1093/nar/gkaa1100
- Välíkangas, T., Suomi, T., and Elo, L. L. (2016). A systematic evaluation of normalization methods in quantitative label-free proteomics. *Brief. Bioinf.* 19:bbw095. doi: 10.1093/bib/bbw095
- Van Eyk, J. E., and Snyder, M. P. (2018). Precision medicine: Role of proteomics in changing clinical management and care. *J. Proteome Res.* 18, 1–6. doi: 10.1021/acs.jproteome.8b00504
- Williams, S. A., Kivimäki, M., Langenberg, C., Hingorani, A. D., Casas, J. P., Bouchard, C., et al. (2019). Plasma protein patterns as comprehensive indicators of health. *Nat. Med.* 25, 1851–1857. doi: 10.1038/s41591-019-0665-2
- Wilson, D. H., Rissin, D. M., Kan, C. W., Fournier, D. R., Piech, T., Campbell, T. G., et al. (2016). The Simoa HD-1 analyzer: A novel fully automated digital immunoassay analyzer with single-molecule sensitivity and multiplexing. *J. Lab. Autom.* 21, 533–547. doi: 10.1177/2211068215589580
- Wingo, A. P., Liu, Y., Gerasimov, E. S., Gockley, J., Logsdon, B. A., Duong, D. M., et al. (2021). Integrating human brain proteomes with genome-wide association data implicates new proteins in Alzheimer's disease pathogenesis. *Nat. Genet.* 53, 143–146. doi: 10.1038/s41588-020-00773-z
- Yeh, C. Y., Adusumilli, R., Kullolli, M., Mallick, P., John, E. M., and Pitteri, S. J. (2017). Assessing biological and technological variability in protein levels measured in pre-diagnostic plasma samples of women with breast cancer. *Biomarker Res.* 5:30. doi: 10.1186/s40364-017-0110-y
- Zaghlool, S. B., Sharma, S., Molnar, M., Matías-García, P. R., Elhadad, M. A., Waldenberger, M., et al. (2021). Revealing the role of the human blood plasma proteome in obesity using genetic drivers. *Nat. Commun.* 12:1279. doi: 10.1038/s41467-021-21542-4
- Zhang, F., Ge, W., Ruan, G., Cai, X., and Guo, T. (2020). Data-independent acquisition mass spectrometry-based proteomics and software tools: A glimpse in 2020. *Proteomics* 20:1900276. doi: 10.1002/pmic.201900276
- Zhong, W., Edfors, F., Gummeson, A., Bergström, G., Fagerberg, L., and Uhlen, M. (2021). Next generation plasma proteome profiling to monitor health and disease. *Nat. Commun.* 12:2493. doi: 10.1038/s41467-021-22767-z

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Correa Rojo, Heylen, Aerts, Thas, Hooyberghs, Ertaylan and Valkenburg. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.