Check for updates

OPEN LETTER

REVISED Foresight in clinical proteomics: current status, ethical considerations, and future perspectives

[version 2; peer review: 3 approved]

Filip Mundt[1], Nicolai J. Wewer Albrechtsen[1,2], Sebastian Porsdam Mann[3], Peter Treit[4], Medini Ghodgaonkar-Steger[4], Martina O'Flaherty[5], Reinout Raijmakers[5], Juan Antonio Vizcaíno[6], Albert J.R. Heck[5], Matthias Mann[1,4]

[1]Novo Nordisk Foundation Center for Protein Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark
[2]Department of Clinical Biochemistry, University Hospital, Bispebjerg Hospital, Bispebjerg, Denmark
[3]Faculty of Law, University of Oxford, Oxford, England, UK
[4]Max Planck Institute of Biochemistry, Proteomics and Signal Transduction, Martinsried, Germany
[5]Biomolecular Mass Spectrometry and Proteomics, Bijvoet Centre for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Padualaan 8, Utrecht 3584 CH, The Netherlands
[6]European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Wellcome Trust Genome Campus, Hinxton, Cambridge, CB10 1SD, UK

v2 **First published:** 18 Apr 2023, **3**:59
https://doi.org/10.12688/openreseurope.15810.1
Latest published: 01 Nov 2023, **3**:59
https://doi.org/10.12688/openreseurope.15810.2

Abstract

With the advent of robust and high-throughput mass spectrometric technologies and bioinformatics tools to analyze large data sets, proteomics has penetrated broadly into basic and translational life sciences research. More than 95% of FDA-approved drugs currently target proteins, and most diagnostic tests are protein-based. The introduction of proteomics to the clinic, for instance to guide patient stratification and treatment, is already ongoing. Importantly, ethical challenges come with this success, which must also be adequately addressed by the proteomics and medical communities. Consortium members of the H2020 European Union-funded proteomics initiative: European Proteomics Infrastructure Consortium-providing access (EPIC-XS) met at the Core Technologies for Life Sciences (CTLS) conference to discuss the emerging role and implementation of proteomics in the clinic. The discussion, involving leaders in the field, focused on the current status, related challenges, and future efforts required to make proteomics a more mainstream technology for translational and clinical research. Here we report on that discussion and provide an expert update concerning the feasibility of clinical proteomics, the ethical implications of generating and analyzing

Open Peer Review

Approval Status ✓ ✓ ✓

	1	2	3
version 2 (revision) 01 Nov 2023	✓ view		
version 1 18 Apr 2023	✓ view	✓ view	✓ view

1. **Nurhan Ozlu**, Koc Universitesi, Istanbul, Turkey
Daniel J. Geiszler, Koc Universitesi, Istanbul, Turkey

2. **Maria Hernandez-Valladares**, University of Granada, Granada, Spain

3. **Kellye A Cupp-Sutton**, The University of Oklahoma, Norman, USA

Any reports and responses or comments on the

large-scale proteomics clinical data, and recommendations to ensure both ethical and effective implementation in real-world applications.

article can be found at the end of the article.

Keywords
Clinical proteomics; clinical research; ethical challenges



This article is included in the Health Sciences gateway.



This article is included in the Clinical Medicine gateway.



This article is included in the Horizon 2020 gateway.



This article is included in the Healthcare and Policy collection.

Corresponding author: Matthias Mann (mmann@biochem.mpg.de)

Author roles: Mundt F: Conceptualization, Writing – Original Draft Preparation; **Albrechtsen NJW**: Conceptualization, Writing – Original Draft Preparation; **Mann SP**: Conceptualization, Writing – Original Draft Preparation; **Treit P**: Conceptualization, Writing – Original Draft Preparation; **Ghodgaonkar-Steger M**: Conceptualization, Writing – Review & Editing; **O'Flaherty M**: Conceptualization, Writing – Review & Editing; **Raijmakers R**: Writing – Review & Editing; **Vizcaíno JA**: Writing – Review & Editing; **Heck AJR**: Writing – Review & Editing; **Mann M**: Conceptualization, Writing – Review & Editing

Competing interests: M. M. is an indirect investor in Evosep Biosystems. All other authors declare no competing interests.

Grant information: This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 823839 (European Proteomics Infrastructure Consortium providing Access [EPIC-XS]). Nicolai J. Wewer Albrechtsen is supported by an NNF Excellence Emerging Investigator Grant – Endocrinology and Metabolism (Application No. NNF19OC0055001), EFSD Future Leader Award (NNF21SA0072746), and DFF Sapere Aude (1052-00003B). Novo Nordisk Foundation Center for Protein Research is supported financially by the Novo Nordisk Foundation (Grant agreement NNF14CC0001). The Netherlands Proteomics Centre (AJRH) acknowledges support from the Netherlands Organization for Scientific Research (NWO) through the X-omics Road Map program (project 184.034.019). JAV wants to acknowledge funding from Wellcome (223745/Z/21/Z) and EMBL core funding.
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2023 Mundt F *et al*. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Mundt F, Albrechtsen NJW, Mann SP *et al.* **Foresight in clinical proteomics: current status, ethical considerations, and future perspectives [version 2; peer review: 3 approved]** Open Research Europe 2023, **3**:59 https://doi.org/10.12688/openreseurope.15810.2

First published: 18 Apr 2023, **3**:59 https://doi.org/10.12688/openreseurope.15810.1

REVISED **Amendments from Version 1**

Following the review comments, we've made several changes to the document, including clarifications and elaborations on certain points, rearranging sections for better flow, adding references for credibility, and improving the overall text quality. These changes make the document more user-friendly, coherent, and reliable.

Any further responses from the reviewers can be found at the end of the article

Introduction

Proteins inherently reflect a more functional or phenotypic view of a (liquid) tissue's status compared to the transcriptome or genome[1]. Clinics routinely use protein-related information for diagnostic and therapeutic decisions as more than 95% of FDA-approved drugs currently target proteins[2], additionally hundreds of FDA-approved protein-based clinical biomarkers are readily detectable in blood alone[3]. The stunning development of robust and high-throughput proteomics techniques and workflows in the last decade has broadened access to proteomics, and in conjunction with ever-advancing bioinformatics tools, the biological and translational research fields are increasingly embracing the proteome. Liquid chromatography (LC) coupled with mass spectrometry (MS), known as LC-MS/MS, is a burgeoning technique renowned for its capacity to quantify tens of thousands of proteins, discern various post-translational modifications[4,5], ascertain the protein content of the extracellular matrix[6,7] (22159717), and elucidate intricate protein-protein interactions[8,9]. This methodology employs the ionization of peptides and their subsequent weighing in a vacuum by directing them towards a mass analyzer[10]. MS-based proteomics represents an exceptional approach for probing the proteome in all its complexities, a feat that remains challenging or near-impossible using other omics methods. Whether a proteome is analyzed by itself or in conjunction with paired nucleic acid (*proteogenomics*), more proteomes than ever are now being generated. Such readily and economically generated proteomes raise hopes for universal wellness or disease tests that could change the practice of medicine. However, along with promises come potential pitfalls. As a community, we have a responsibility to discuss and address ethical considerations in implementing MS-based proteomics in the clinic.

In the present manuscript, we aim to summarize the fruitful discussions that started at the online Core Technologies for Life Science Conference (CTLS; https://ctls-org.eu/) conference in September 2021, with a focus on clinical implementation of MS-based proteomics. We also discuss ethical implications and important considerations in acquiring proteomics data from patients, including 'incidental findings,' as well as the possibility of re-identifying (anonymized) patient samples. We conclude by providing recommendations for further discussion of the issues raised herein.

European Proteomics Infrastructure Consortium - Providing Access

This manuscript is part of the efforts of European Proteomics Infrastructure Consortium - Providing Access (EPIC-XS), which is a project funded by the European Union to provide proteomics expertise and mass spectrometry technology to researchers within the life science arena (https://epic-xs.eu/). It brings together eighteen institutes, spread across fifteen European countries, with the objective of providing more than 2,400 days of access to high-end proteomics technologies. The EPIC-XS consortium tackles fundamental aspects of the consolidation and enhancement of proteomics technologies and skills across Europe. The EPIC-XS infrastructure contributes to strengthening Europe's leading role in proteomics by ensuring access of a broad range of researchers to prominent European proteomics facilities, on the premise that research in this field can yield information spanning the whole life science domain, including developments towards potential disease biomarkers. EPIC-XS has an independently appointed ethics advisor to help the consortium ensure the compliance of the project with ethical standards (national and EU) and guidelines of the H2020 program.

Proteomics in the clinic

A plurality of clinical diagnostics utilizes protein-based biomarkers and at large hospitals, there may be as many as three to five million analyses of protein-based biomarkers per year[3]. However, *proteomics* is still not used in the diagnostic units or in laboratory medicine, for reasons that include perceived technological immaturity, a feared lack of robust and accurate measurements, as well as an apparent high cost. These, in turn, mirror problems seen in other areas of omics research applied to translational medicine, such as a lack of external validation cohorts, insufficient statistical power, and the use of inadequate comparator groups.

In addition, there are additional perceived issues that are specific to clinical proteomic studies. Whereas traditional biomarker-based studies aim to use material from well-controlled standardized clinical trials and from standard operating procedures (SOP), clinical proteomics must be able to handle conditions where the quality of biomaterial sampling and pre-analytical requirements are variable, and sampling procedures are often optimized for measuring devices other than a mass spectrometer. Optimized SOPs may rarely be used for collecting patient material in these settings and therefore, several developments must occur before MS-based proteomics can be successfully implemented in clinical diagnostics. In laboratory medicine, the goal is to provide results that can guide all aspects of patient care, which require robust and reproducible data, even for serial patient samples taken at only a few minutes' intervals. Many published studies report potential biomarkers of multiple diseases, yet very few, if any, reach clinical application (although proteomics is not alone in this)[11]. For clinical studies, large sample collections and cohorts are needed, ideally taken by longitudinal sampling. There are several promising

developments in this regard, including newer, faster, MS hardware that allows for the analysis of larger cohorts (in the thousands), statistical tools that help to evaluate variance[12] and improve power calculations[13], as well as meta-analyses of public proteomics data sets[14]. These developments are complemented by ongoing efforts to increase hardware robustness for large cohorts[15,16], and by targeted MS panels that can be used with relative ease on hundreds, or even thousands of samples for validation efforts[17,18]. In the past, cohorts that serendipitously, or purposefully, have been used for cross validation have been analyzed in different proteomic laboratories[4,19]. Additionally, the use of exchanging reference samples can beneficiary align proteomic laboratories[20]. However, the more ubiquitous implementation of such practices is not anything we see in the field at the moment.

Clinical proteomics is likely to play a significant role in clinics by providing a more specific, quantitatively accurate and possibly cheaper platform for diagnostics, as compared to some current gold standards. For example, diagnoses of multiple myeloma or hemoglobinopathies are currently based on gel electrophoresis and liquid chromatography[21], which can take up to two weeks. In our experience, MS-based proteomics can do this in hours and, unlike visual detection (by eye)[22–24], is compatible with machine learning approaches, potentially reducing (human) error. In addition, proteomics may directly provide answers that are normally resolved through subsequent genetic testing, for example for hemoglobinopathies

The potential and difficulties of clinical applications of proteomics are well illustrated by our experience at two major Danish hospitals (firstly, setting up our MS-based proteomics at Rigshospitalet and then moving this setup to the Bispebjerg Hospital). Since 2020, we have established an MS-based proteomics setup within the hospital, which includes an EvoSep One liquid chromatography (EvoSep), and an Orbitrap Exploris 480™ mass spectrometer (Thermo Fisher Scientific). In many cases, we have found that MS-based proteomics readily outperforms current clinical diagnostic praxis as a diagnostic tool. Examples include the aforementioned diagnostics of multiple myeloma and hemoglobinopathies, as well as the ability to discriminate between bacterial, viral, and trauma-induced host-related responses. For instance, a simple blood sample can be used to identify patients at high risk of the tick-borne disease borreliosis, as opposed to risky and invasive spinal tap procedure. While it is uncertain how these efforts will be integrated into broad routine clinical practice, physicians in the capital region of Denmark are currently able to request plasma proteomics for routine diagnostic analysis. We are limited by a lack of standardization protocols, or of validation reports in accordance with the standards set forth by the national accreditation institutions as well as by a lack of a universally agreed-upon human proteome standard or reference intervals. Although these issues challenge contemporary MS-based clinical applications, our experience suggests that they are manageable.

Mass spectrometry-based proteomics has created new insights in disease understanding. Landscaping studies have connected somatic mutations to breast cancer signaling[25], highlighted subsets of tumors for immune therapy[26–28], identified ALK-fusion diagnostic markers and targets in lung cancer[26], and stratified new patient subgroups of medulloblastoma that are invisible via the study of the genome and transcriptome[4,6].

As a demonstration of the current robustness and reproducibility of quantitative proteomics, three different laboratories (two of which are a partner in EPIC-XS), using different patient cohorts and different methods, independently defined very similar serum proteome signatures that predict disease severity and mortality in COVID-19 patients[29–31]. There is considerable work ongoing aiming to standardize these workflows to uphold the International Organization for Standardization (ISO) standard for quality of medical devices (ISO 13485[29,32]). Also in the plasma proteome, post-translational modifications are of fundamental importance, as exemplified by the use of glycoproteins in plasma to improve prediction of patient outcomes[33].

In another example, the EPIC-XS center in Copenhagen analyzed plasma samples from nearly 600 patients with alcohol-related liver disease. We were able to predict future liver-related events and all-cause mortality with a Harrell's C-index of 0.90 and 0.79, respectively. Importantly, the diagnostic model performance was reproduced in an independent validation cohort reproduced[34], laying the foundation for routine MS-based liver disease testing. Additionally, in a first-of-its-kind publication, we also described a rapid proteomics workflow to analyze and predict new treatment modalities for a patient with terminal cancer, with the outcome approved by the cancer board[35]. Such analyses might well become standard in future settings of precision oncology and medicine[36]. We hope to see an era of improved diagnostics, and crucially, improved patient care facilitated by increased use of MS-based proteomics in the clinic.

Clinical proteomics in the era of the EU General Data Protection Regulation

The General Data Protection Regulation (GDPR) is an EU law governing data protection and privacy. The GDPR is generally applicable and thus also binding for all research carried out partially or wholly in the European Union and the European Economic Area and has served as an inspiration for data privacy legislation globally. It is therefore also relevant for clinical proteomics. Though proteomics information is intended to aid doctors in their decision-making processes, from diagnostics to prognostics and treatment, it can potentially be put to other uses, including the re-identification of individuals in a proteomic dataset. Identifiable information is information, such as telephone or social security numbers, birthdates, and addresses that might be used to identify a unique individual. By contrast, re-identifiable information is information that enables one to match one sample to another in a dataset. Re-identifiability may be achievable from different data types that can be generated in proteomics experiments, such as protein expression patterns, specific patterns in protein post-translational modifications, and from allotype-induced amino-acid sequence

variation (including single amino acid variants)[37]. Presently, sequence information from proteomics data is less likely to re-identify an individual than genomic or transcriptomic sequence data[38,39]. The re-identification of an individual may enable the derivation of sensitive information about that individual or other infringements of privacy. More pertinently, such data typically contains several incidental findings; the question of whether to return these findings hinges on the medical actionability of the finding (*e.g.,* biomarkers), as well as patient preference. This is true even for retrospective proteomics studies (*i.e.*, reanalysis of past data, which is available in the public domain). It should be noted that improvements in proteomics data analysis and computation may lead to data becoming re-identifiable (or identifiable) in a few years in the future, even if it was not originally the case when the data was originally generated and made available in a public repository. This needs to be taken into account when developing guidelines.

Re-identifiable information has the potential to become identifiable when combined with metadata that is itself identifiable. In one such effort to reanalyze proteomics data from a previously published longitudinal weight loss study of 1,500 individuals over 14 months, we found plasma proteomes uniquely re-identifiable within the cohort based on individual-specific protein expression levels as well as allotype-specific variant peptides[40,41]. Additionally, vitamin D-binding proteins vary significantly based on ethnic background and can therefore be used to derive the ethnicity of study participants, which is also readily apparent from the distribution of coding single nucleotide variants (cSNPs)[42]. Levels of pregnancy-zone protein could be used to determine whether an individual is pregnant or post-menopausal, and various hormone-related proteins can also be used to identify biological sex. These findings illustrate the potential for deriving sensitive information about individuals from their proteomes, provided sufficient metadata is available. These conclusions are ethically significant since such information could in theory be used to disadvantage or discriminate against individuals.

In the same weight loss study, the APOE allotype status could be determined from the proteomic data. APOE alleles are associated with different cardiovascular and Alzheimer's disease risks[41]. The APOE4 allele strongly increases the risk of Alzheimer's disease, whereas the APOE2 allele is associated with increased cholesterol levels and cardiovascular pathologies. Similarly, the presence of elevated levels of glycated forms of hemoglobin protein can indicate (pre-) diabetic pathology. Likewise, the levels of proteins associated with inflammation, lipid metabolism, and various hormones, could all point to different likely health and disease states, which is indeed the primary goal of plasma proteome profiling. Complicating the question of what should be done with such information is the fact that only some of these (e.g., glycated hemoglobin and APOE2 in the examples above) are actionable, meaning that this knowledge could benefit patient's health and wellbeing, whereas some (*e.g*., APOE4) are currently unactionable and may instead lead to distress, when this information is communicated to the patient; actionable information is information that, if returned to individuals, can be acted on to their benefit.

In the Utrecht EPIC-XS site, alternative plasma proteomics approaches are being explored to monitor immunoglobulin repertoires[43] and proteoform profiles (*i.e*., the full compendium of different forms of a protein, including protein species that contain different sets of post-translational modifications) or serum glycoproteins[33]. In these studies, small liquid biopsies (blood, human milk) are taken from individual patients, whereafter either specific classes of immunoglobulins or selected plasma glycoproteins are extracted, which are subsequently analyzed and identified by MS. These preliminary studies revealed that such traits of the plasma proteome can be rather unique for each donor and thus potentially even more tractable than data obtained by protein expression profiling described above. These traits are not directly traceable to the genome sequences of the donor and thus represent more unique features of the donor's proteome. It needs to be seen whether such data also provides actionable results. However, incidentally, our analyses have already led to the discovery of a B-cell malignancy in one of the donors, which led the doctors to change therapy. Although these approaches were performed using 'top-down proteomics', a still less mature technology than the more standard 'bottom-up' expression proteomics pipelines described above, they may point to additional issues that need to be addressed from an ethical angle.

Even with these risks, sharing clinical proteomics data, with consent from the patients, is the goal for most proteomics researchers. There are already standardized community guidelines for data submission and dissemination of proteomics data in the public domain[44,45], which were developed under the umbrella of the ProteomeXchange Consortium, and, are tied to the requirements from either funding agencies, scientific journals - or both. The current community standard is to upload MS raw files and the processed results to the PRoteomics IDEntifications database (PRIDE, at the European Bioinformatics Institute [EMBL-EBI]), which is the leading resource within ProteomeXchange[44,45]. The wide availability of public data also tends to increase the quality of data and clinical research because of more transparent procedures and possibilities for proteomics data re-use including meta-analyses, among many other applications [28118949][46]. The overall goal is to make the data findable, accessible, interoperable, and reusable after the (FAIR) principles[47].

In the EU, the GDPR has strengthened privacy protection but, as a consequence this has made it more difficult to enable sharing of data between academic research institutes, which was not necessarily the original intention in the medical context. The GDPR regulations are complemented by other regional or national regulations such as the Health Insurance Portability and Accountability Act (HIPAA) in the United States of America, and the Act on the Protection of Personal Information (APPI) in Japan. The current state of the art with respect to data management practices of sensitive human data in proteomics has been summarized recently[38].

To illustrate the unintended consequences of the GDPR, we have experienced a direct example of a clash between the proteomics community guidelines and GDPR as part of a large

national collaboration in Denmark where we analyzed plasma samples from patients with alcohol-related liver disease[34]. In this study, we planned to follow the ProteomeXchange guidelines, by uploading the MS raw files, as well as standard clinical data (*e.g.*, age and body–mass index), to the PRIDE database. However, the hospital's compliance officers concluded that this data qualified as personal clinical data, the publication of which required specific permission that that not available at the time. After considerable delay (almost a year), we resolved the issue by providing the data in a "controlled access" manner, meaning that a user wanting to get access to the data has to apply for access at the Danish Data Protection Agency (DDPA) and be approved by an ethics committee for the Region of Southern Denmark, which is a lengthy process. ProteomeXchange resources including PRIDE were originally set up to contain fully open data (similarly to resources such as GEO[48] and ArrayExpress[49] in the transcriptomics field). PRIDE does not currently offer controlled access. Other resources such as the European Genome-phenome Archive (EGA) do, but they were set up for DNA/RNA sequencing data and their data model is tailored for this functionality. PRIDE has started to work in a controlled access repository for proteomics data. This will still take some time to be in production, but it is hoped to help researchers to solve the issues that they are facing in this context, and at the same type provide tailored support for proteomics data. Obtaining controlled access data is a generally lengthy and cumbersome process—not an issue exclusive to proteomics. This will unfortunately hinder many researchers from accessing and ultimately using the data and is thus arguably itself of ethical concern. It might also lead to a bias in who will use it, with smaller institutions lacking funding and resources to acquire access, thus favoring large, well-funded institutions.

How is ethical analysis relevant to a forthcoming era of clinical proteomics?

We recently surveyed the recent discussion of ethics in clinical proteomics and identified four reasons to engage seriously with the topic[50]. Among the main findings were, first, that ethics is also about finding ways to increase the beneficial impact of the work that we do, rather than only an obstacle to be overcome. Second, experience from related fields, especially genomics, shows that early discussion of these issues can be beneficial, because informed self-regulation has numerous advantages over the alternative of having external legislation being imposed from those not within the field of proteomics. Third, we can learn from the way similar issues have been handled in genomics and other fields; and fourth, despite this similarity, it is important to consider as early as possible which issues might be different and unique to the clinical proteomics context, when compared with nucleotide sequencing data.

The discipline of bioethics provides a framework useful for identifying and discussing ethical issues emerging by innovations in biotechnology and medicine. The bioethical methodology involves the application of normative principles gleaned from ethical traditions to scientific and medical contexts. These principles, briefly, are: (i) beneficence, which stresses the ethical value of benefiting people; (ii) non-maleficence, which conversely concerns the ethical imperative of not harming persons; (iii) justice, or the value of treating people fairly and equitably; and (iv) autonomy, requiring respect for the free and informed decisions of individuals. Although these principles do not themselves provide a method for determining what ought to be done, they do provide a framework useful for clarifying underlying issues of ethical importance.

Using these bioethical principles as operational definitions of 'ethical issues' (an issue is ethical if it is identified as such or obviously relates to these principles) we performed a systematic review of ethical issues already discussed in the clinical proteomics literature. By employing qualitative analysis, we found 40 ethical issues across 16 included studies and grouped these into 10 ethical themes, varying from the importance of standards and quality control to the need for international discussion and development of guidelines on ethical issues in clinical proteomics[50]. The themes also included incidental findings, re-identifiability, and the potential for discovery of sensitive information previously identified as ethically important topics, as described above[40].

Most of these issues had already been discussed in the context of clinical genomics. As a result, there is a high degree of consensus on important topics in this field. The most pertinent lessons from genomics concern the handling of sensitive data and incidental findings, as well as the importance of early and serious discussion of ethical issues to avert subsequent externally imposed regulations. For example, the distinction between actionable and unactionable information is of great ethical and practical significance. Reviews in the literature generally concluded that actionable information ought to be returned to the individual or their health care provider, whereas unactionable information should not[51,52]. These reviews also argued that individuals should be informed of the likelihood of such findings at the point of consent and that individuals' preferences as expressed during the consent process regarding the return of findings should be respected. The literature on the return of incidental findings in genomics also pointed out that attempts at determining the actionability and health relevance of findings are hampered by the high prevalence of variants of uncertain significance. There are additional issues concerning whether incidental findings should be routinely incorporated in health registries and other scientific databases as well as what should be done with incidental findings arising from the reanalysis of old data. Most of these topics will be relevant for proteomics and it is clear in our view that the field can certainly learn a great deal from that discussion.

One area of divergence between genomics and proteomics is the ability of the latter to capture phenotypic information. Proteomic information can be used to provide actionable health information to individuals and to advance biomedical understanding. Notably, this can be done whatever the health state, and healthy individuals could also receive health-relevant information, which they could incorporate into their daily lifestyles

if they wished to do so. However, these profiles should be treated with caution as they could contain all kinds of potentially sensitive information, much of which will be unrelated to the reasons why people chose to be profiled. What, then, should be done with this extra information? A simple method of avoiding these ethical questions is simply not to look for incidental findings or to anonymize data sets completely. However, failing to analyze or share health-related information can result in missed opportunities to improve the individual's and other people's lives.

Serious, sustained, and early treatment of these and other ethical issues are likely to yield sustained benefits by informing the development of guidelines on ethical issues by the clinical proteomics community. Self-regulation in clinical proteomics is not only important because it directly involves the scientists' expertise in the field but also because guidelines developed through internal consensus are more likely to be perceived as legitimate and therefore be effective.

Conclusion

The purpose of the clinical proteomics round table discussion that constitutes the foundation of this paper was to start a comprehensive conversation aimed at identifying key challenges and proposing solutions to maintain the trustworthiness and effectiveness of clinical proteomics. Efforts towards facilitating these goals will help serve in incorporating clinical proteomics technologies in all aspects of patient management. Though many critical questions were raised during our discussion, and while it was clear to all that there is a long and potentially difficult road ahead, there was nevertheless broad consensus on the bright outlook for clinical proteomics as a strong asset to clinicians. By communicating the benefits, clarifying SOPs, and initiating discussions at international scientific meetings, we hope to discuss and eventually resolve some of the ethical and regulatory issues that we are currently hampered by. In particular, we propose:

A natural follow-up of EPIC-XS, *i.e.*, a strong joint EU proteomics infrastructure community that should work toward setting up SOPs for MS-based clinical proteomics workflows and methods, in accordance with the standards set forth by the national medical accreditation institutions in various EU countries. The aim should be not only to increase throughput and reduce costs but also to work according to the FAIR principles whenever possible given the GDPR guidelines and to generate robust and reproducible data across different centers.

Addressing the need to develop tailored bioinformatics infrastructure to handle and manage sensitive human (clinical) proteomics data. The most urgent development is the availability of controlled access functionality in public proteomics data repositories. Controlled access functionality is already available for DNA/RNA sequencing data in resources such as the EGA, but their data model was originally developed for DNA/RNA sequencing data and cannot appropriately represent proteomics datasets. Although the first step would be the availability of a central controlled access data infrastructure, there will be a need to build on existing infrastructures that are amenable to cross-border data sharing without the data actually crossing borders e.g. the European Life-sciences Infrastructure for biological Information (ELIXIR) (https://elixir-europe.org/) national nodes (*e.g.* a federated infrastructure). This could be achieved under an umbrella structure and involve partners such as EMBL-EBI, the Biobanking and BioMolecular resources Research Infrastructure-European Research Infrastructure Consortium (BBMRI-ERIC), or European Strategy Forum on Research Infrastructures (ESFRI), and/or ELIXIR, among others.

Data access committees are key since they provide access to researchers for human-sensitive datasets in the public domain. We need to strengthen our dialogues with clinicians, patient interest groups, and data access committees about the particularities of proteomics data which may generally be less risky for patient re-identification, when compared with nucleotide sequencing information. There is also a need for further research studies in this area to provide as much accurate and useful information as possible to take properly informed decisions.

Developing clear guidelines on how to handle human-sensitive clinical proteomics data. Importantly, they should address how to handle "incidental findings", both actionable and unactionable, when searching or re-searching proteomics data. Furthermore, guidelines will need to be developed concerning the submission and dissemination of such data across public resources. This can best be accomplished through collaborative multinational efforts involving clinicians, patient interest groups, bioethicists, data protection specialists, and proteomics experts, underscoring the significance of global cooperation.

These are all mid- to long-term objectives on which our sibling platforms in genomics and transcriptomics have been working for years. From our discussions, it also became clear that there needs to be a more concerted effort not only in Europe but ideally worldwide in formulating a strategy wherein clinical proteomics can achieve sustainability. These and other issues could be addressed at one of the large annual conferences attended by a significant portion of the community, such as the Human Proteome Organization (HUPO) world congress.

Proactively addressing the above-mentioned challenges in a cohesive way would lead to the development of valuable and appropriate guidelines for the clinical proteomics community. This would be preferable to having external governance impose restrictions on the field. We believe that this is a very important goal to maximize the benefit of the coming revolution in clinical proteomics for the individual and for society.

Data and software availability

No data are associated with this article.

Author contributions

FM, NJWA, SPM, PT, MGS, MOF and MM were involved in Conceptualization, meeting organization and leading the round table discussion at the CTLS conference. FM, NJWA, SPM, PT, wrote the Original Draft Preparation of the manuscript. MGS, MOF, MM, RR, JAV and AJRH contributed significantly to the writing, through Review and Editing of the manuscript.

Acknowledgments

We acknowledge valuable input and discussions from several other EPIC XS consortium members, including Richard Scheltema*, Eduard Sabido§, Tiziana Bonaldi#, and Mariette Matondo¶. We would furthermore like to extend our gratitude to Lili Niu¤ for constructive discussions around access to clinical proteomics data.

* Utrecht University, Padualaan 8, NL-3584 CH Utrecht, The Netherlands

§ Center for Genomic Regulation, Barcelona, Spain

European Institute of Oncology IRCCS, Department of Experimental Oncology, IEO, Milan, Italy

¶ Institut Pasteur, Center National Research Science, Paris, France

¤ Novo Nordisk Foundation Center for Protein Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen Denmark

References

1. Wang J, Ma Z, Carr SA, *et al.*: **Proteome Profiling Outperforms Transcriptome Profiling for Coexpression Based Gene Function Prediction.** *Mol Cell Proteomics.* 2017; **16**(1): 121–134.
PubMed Abstract | Publisher Full Text | Free Full Text

2. Santos R, Ursu O, Gaulton A, *et al.*: **A comprehensive map of molecular drug targets.** *Nat Rev Drug Discov.* 2017; **16**(1): 19–34.
PubMed Abstract | Publisher Full Text | Free Full Text

3. Geyer PE, Holdt LM, Teupser D, *et al.*: **Revisiting biomarker discovery by plasma proteomics.** *Mol Syst Biol.* 2017; **13**(9): 942.
PubMed Abstract | Publisher Full Text | Free Full Text

4. Archer TC, Ehrenberger T, Mundt F, *et al.*: **Proteomics, Post-translational Modifications, and Integrative Analyses Reveal Molecular Heterogeneity within Medulloblastoma Subgroups.** *Cancer Cell.* 2018; **34**(3): 396–410.e8.
PubMed Abstract | Publisher Full Text | Free Full Text

5. Ochoa D, Jonikas M, Lawrence RT, *et al.*: **An atlas of human kinase regulation.** *Mol Syst Biol.* 2016; **12**(12): 888.
PubMed Abstract | Publisher Full Text | Free Full Text

6. Eckert MA, Coscia F, Chryplewicz A, *et al.*: **Proteomics reveals NNMT as a master metabolic regulator of cancer-associated fibroblasts.** *Nature.* 2019; **569**(7758): 723–728.
PubMed Abstract | Publisher Full Text | Free Full Text

7. Naba A, Clauser KR, Hoersch S, *et al.*: **The matrisome: *in silico* definition and *in vivo* characterization by proteomics of normal and tumor extracellular matrices.** *Mol Cell Proteomics.* 2012; **11**(4): M111 014647.
PubMed Abstract | Publisher Full Text | Free Full Text

8. Prabhu N, Dai L, Nordlund P: **CETSA in integrated proteomics studies of cellular processes.** *Curr Opin Chem Biol.* 2020; **54**: 54–62.
PubMed Abstract | Publisher Full Text

9. Titeca K, Lemmens I, Tavernier J, *et al.*: **Discovering cellular protein-protein interactions: Technological strategies and opportunities.** *Mass Spectrom Rev.* 2019; **38**(1): 79–111.
PubMed Abstract | Publisher Full Text

10. Aebersold R, Mann M: **Mass spectrometry-based proteomics.** *Nature.* 2003; **422**(6928): 198–207.
PubMed Abstract | Publisher Full Text

11. Mischak H, Ioannidis JPA, Argiles A, *et al.*: **Implementation of proteomic biomarkers: making it work.** *Eur J Clin Invest.* 2012; **42**(9): 1027–36.
PubMed Abstract | Publisher Full Text | Free Full Text

12. Oberg AL, Vitek O: **Statistical design of quantitative mass spectrometry-based proteomic experiments.** *J Proteome Res.* 2009; **8**(5): 2144–56.
PubMed Abstract | Publisher Full Text

13. Ding LJ, Schlüter HM, Szucs MJ, *et al.*: **Comparison of Statistical Tests and Power Analysis for Phosphoproteomics Data.** *J Proteome Res.* 2020; **19**(2): 572–582.
PubMed Abstract | Publisher Full Text | Free Full Text

14. Dogu E, Taheri SM, Olivella R, *et al.*: **MSstatsQC 2.0: R/Bioconductor Package for Statistical Quality Control of Mass Spectrometry-Based Proteomics Experiments.** *J Proteome Res.* 2019; **18**(2): 678–686.
PubMed Abstract | Publisher Full Text

15. Mc Ardle A, Binek A, Moradian A, *et al.*: **Standardized Workflow for Precise Mid- and High-Throughput Proteomics of Blood Biofluids.** *Clin Chem.* 2022; **68**(3): 450–460.
PubMed Abstract | Publisher Full Text

16. Bader JM, Geyer PE, Müller JB, *et al.*: **Proteome profiling in cerebrospinal fluid reveals novel biomarkers of Alzheimer's disease.** *Mol Syst Biol.* 2020; **16**(6): e9356.
PubMed Abstract | Publisher Full Text | Free Full Text

17. van Bentum M, Selbach M: **An Introduction to Advanced Targeted Acquisition Methods.** *Mol Cell Proteomics.* 2021; **20**: 100165.
PubMed Abstract | Publisher Full Text | Free Full Text

18. Kennedy JJ, Whiteaker JR, Ivey RG, *et al.*: **Internal Standard Triggered-Parallel Reaction Monitoring Mass Spectrometry Enables Multiplexed Quantification of Candidate Biomarkers in Plasma.** *Anal Chem.* 2022; **94**(27): 9540–9547.
PubMed Abstract | Publisher Full Text | Free Full Text

19. Forget A, Martignetti L, Puget S, *et al.*: **Aberrant ERBB4-SRC Signaling as a Hallmark of Group 4 Medulloblastoma Revealed by Integrative Phosphoproteomic Profiling.** *Cancer Cell.* 2018; **34**(3): 379–395.e7.
PubMed Abstract | Publisher Full Text

20. Mertins P, Tang LC, Krug K, *et al.*: **Reproducible workflow for multiplexed deep-scale proteome and phosphoproteome analysis of tumor tissues by liquid chromatography-mass spectrometry.** *Nat Protoc.* 2018; **13**(7): 1632–1661.
PubMed Abstract | Publisher Full Text | Free Full Text

21. Willrich MA, Katzmann JA: **Laboratory testing requirements for diagnosis and follow-up of multiple myeloma and related plasma cell dyscrasias.** *Clin Chem Lab Med.* 2016; **54**(6): 907–19.
PubMed Abstract | Publisher Full Text

22. Mallia CP, Nancekivell AF: **Fulminant virus hepatitis in late pregnancy.** *Ann Trop Med Parasitol.* 1982; **76**(2): 143–6.
PubMed Abstract | Publisher Full Text

23. Cho SY, Kim Y, Lee A, *et al.*: **Three Cases Showing False Results in the Detection of Monoclonal Components Using Capillary Electrophoresis.** *Laboratory Medicine.* 2011; **42**(10): 602–606.
Publisher Full Text

24. Šegulja D, Šparakl T, Rogić D: **Applied capillary electrophoresis system affects screening for monoclonal gammopathy in serum: verification study of two eight-capillary systems.** *Scand J Clin Lab Invest.* 2022; **82**(2): 85–89.
PubMed Abstract | Publisher Full Text

25. Mertins P, Mani DR, Ruggles KV, *et al.*: **Proteogenomics connects somatic mutations to signalling in breast cancer.** *Nature.* 2016; **534**(7605): 55–62.
PubMed Abstract | Publisher Full Text | Free Full Text

26. Gillette MA, Satpathy S, Cao S, *et al.*: **Proteogenomic Characterization Reveals Therapeutic Vulnerabilities in Lung Adenocarcinoma.** *Cell.* 2020; **182**(1): 200–225.e35.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

27. Johansson HJ, Socciarelli F, Vacanti NM, *et al.*: **Breast cancer quantitative proteome and proteogenomic landscape.** *Nat Commun.* 2019; **10**(1): 1600.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

28. Krug K, Jaehnig EJ, Satpathy S, *et al.*: **Proteogenomic Landscape of Breast Cancer Tumorigenesis and Targeted Therapy.** *Cell.* 2020; **183**(5): 1436–1456.e31.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

29. Demichev V, Tober-Lau P, Lemke O, *et al.*: **A time-resolved proteomic and prognostic map of COVID-19.** *Cell Syst.* 2021; **12**(8): 780–794.e7.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

30. Geyer PE, Arend FM, Doll S, *et al.*: **High-resolution serum proteome trajectories in COVID-19 reveal patient-specific seroconversion.** *EMBO Mol Med.* 2021; **13**(8): e14167.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

31. Völlmy F, van den Toorn H, Chiozzi RZ, *et al.*: **A serum proteome signature to predict mortality in severe COVID-19 patients.** *Life Sci Alliance.* 2021; **4**(9): e202101099.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

32. Messner CB, Demichev V, Wendisch D, *et al.*: **Ultra-High-Throughput Clinical Proteomics Reveals Classifiers of COVID-19 Infection.** *Cell Syst.* 2020; **11**(1): 11–24.e4.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

33. Čaval T, Lin YH, Varkila M, *et al.*: **Glycoproteoform Profiles of Individual Patients' Plasma Alpha-1-Antichymotrypsin are Unique and Extensively Remodeled Following a Septic Episode.** *Front Immunol.* 2021; **11**: 608466.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

34. Niu L, Thiele M, Geyer PE, *et al.*: **Noninvasive proteomic biomarkers for alcohol-related liver disease.** *Nat Med.* 2022; **28**(6): 1277–1287.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

35. Doll S, Kriegmair MC, Santos A, *et al.*: **Rapid proteomic analysis for solid tumors reveals LSD1 as a drug target in an end-stage cancer patient.** *Mol Oncol.* 2018; **12**(8): 1296–1307.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

36. Doll S, Gnad F, Mann M: **The Case for Proteomics and Phospho-Proteomics in Personalized Cancer Medicine.** *Proteomics Clin Appl.* 2019; **13**(2): e1800113.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

37. Fierro-Monti I, Wright JC, Choudhary JS, *et al.*: **Identifying individuals using proteomics: are we there yet?** *Front Mol Biosci.* 2022; **9**: 1062031.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

38. Bandeira N, Deutsch EW, Kohlbacher O, *et al.*: **Data Management of Sensitive Human Proteomics Data: Current Practices, Recommendations, and Perspectives for the Future.** *Mol Cell Proteomics.* 2021; **20**: 100071.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

39. Keane TM, O'Donovan C, Vizcaino JA: **The growing need for controlled data access models in clinical proteomics and metabolomics.** *Nat Commun.* 2021; **12**(1): 5787.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

40. Geyer PE, Mann SP, Treit PV, *et al.*: **Plasma proteomes can be reidentifiable and potentially contain personally sensitive and incidental findings.** *Mol Cell Proteomics.* 2021; **20**: 100035.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

41. Geyer PE, Albrechtsen NJW, Tyanova S, *et al.*: **Proteomics reveals the effects of sustained weight loss on the human plasma proteome.** *Mol Syst Biol.* 2016; **12**(12): 901.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

42. Schandorff S, Olsen JV, Bunkenborg J, *et al.*: **A mass spectrometry-friendly database for cSNP identification.** *Nat Methods.* 2007; **4**(6): 465–6.
PubMed Abstract | **Publisher Full Text**

43. Bondt A, Hoek M, Tamara S, *et al.*: **Human plasma IgG1 repertoires are simple, unique, and dynamic.** *Cell Syst.* 2021; **12**(12): 1131–1143.e5.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

44. Deutsch EW, Bandeira N, Sharma V, *et al.*: **The ProteomeXchange consortium in 2020: enabling 'big data' approaches in proteomics.** *Nucleic Acids Res.* 2020; **48**(D1): D1145–D1152.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

45. Perez-Riverol Y, Bai J, Bandla C, *et al.*: **The PRIDE database resources in 2022: a hub for mass spectrometry-based proteomics evidences.** *Nucleic Acids Res.* 2022; **50**(D1): D543–D552.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

46. Martens L, Vizcaíno JA: **A Golden Age for Working with Public Proteomics Data.** *Trends Biochem Sci.* 2017; **42**(5): 333–341.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

47. Caufield JH, Fu J, Wang D, *et al.*: **A Second Look at FAIR in Proteomic Investigations.** *J Proteome Res.* 2021; **20**(5): 2182–2186.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

48. Barrett T, Wilhite SE, Ledoux P, *et al.*: **NCBI GEO: archive for functional genomics data sets--update.** *Nucleic Acids Res.* 2013; **41**(Database issue): D991–5.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

49. Sarkans U, Füllgrabe A, Ali A, *et al.*: **From ArrayExpress to BioStudies.** *Nucleic Acids Res.* 2021; **49**(D1): D1502–D1506.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

50. Mann SP, Treit PV, Geyer PE, *et al.*: **Ethical principles, opportunities and constraints in clinical proteomics.** *Mol Cell Proteomics.* 2021; mcp.RA120.002435.
PubMed Abstract

51. Daly MB: **Navigating the Intersection between Genomic Research and Clinical Practice.** *Cancer Prev Res (Phila).* 2020; **13**(3): 219–222.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

52. Saelaert M, Mertes H, De Baere E, *et al.*: **Incidental or secondary findings: an integrative and patient-inclusive approach to the current debate.** *Eur J Hum Genet.* 2018; **26**(10): 1424–1431.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Open Peer Review

Current Peer Review Status: ✔ ✔ ✔

Version 2

Reviewer Report 02 November 2023

https://doi.org/10.21956/openreseurope.18137.r35993

© **2023 Ozlu N et al.** This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

✔ **Nurhan Ozlu**
Department of Molecular Biology and Genetics, Koc Universitesi, Istanbul, İstanbul, Turkey
Daniel J. Geiszler
Department of Molecular Biology and Genetics, Research Center for Translational Medicine, Koc Universitesi, Istanbul, İstanbul, Turkey

The revised version addressed all our comments.

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Proteomics, computational proteomics, cell biology

We confirm that we have read this submission and believe that we have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Version 1

Reviewer Report 14 August 2023

https://doi.org/10.21956/openreseurope.17080.r34062

© **2023 Cupp-Sutton K.** This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

✔ **Kellye A Cupp-Sutton**
[1] The University of Oklahoma, Norman, Oklahoma, USA
[2] The University of Oklahoma, Norman, Oklahoma, USA

This open letter by Mundt *et al.* discusses the various current and potential application of proteomics methods for clinical proteomics. The authors discuss in detail the ethical implications of untargeted and targeted proteomics on anonymized clinical samples. The letter is well written and insightful with in depth analysis of many aspects presented for ethical consideration by the readers. The examples presented for how the ethical considerations have already become an issue in current research serves to further demonstrate the urgency of the discussion. Moreover, the authors present several actionable suggestions for how to move the application of clinical proteomics forward.

I strongly recommend approval of this manuscript.

Is the rationale for the Open Letter provided in sufficient detail? (Please consider whether existing challenges in the field are outlined clearly and whether the purpose of the letter is explained)
Yes

Does the article adequately reference differing views and opinions?
Yes

Are all factual statements correct, and are statements and arguments made adequately supported by citations?
Yes

Is the Open Letter written in accessible language? (Please consider whether all subject-specific terms, concepts and abbreviations are explained)
Yes

Where applicable, are recommendations and next steps explained clearly for others to follow? (Please consider whether others in the research community would be able to implement guidelines or recommendations and/or constructively engage in the debate)
Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Top-down proteomics, analytical biochemistry

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Reviewer Report 07 August 2023

https://doi.org/10.21956/openreseurope.17080.r34055

© 2023 Hernandez-Valladares M. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Maria Hernandez-Valladares
[1] University of Granada, Granada, Spain
[2] University of Granada, Granada, Spain

This Open Letter introduces to the reader some of the challenges we are facing in clinical proteomics. Although most of them are mentioned, the authors mainly focused on very important ethical considerations. I should say that I have missed some more discussion and future perspectives on some other crucial points such as how we could overcome the lack of external validation cohorts and exchange samples, and how to improve the use of more proper comparator groups.

The authors describe some examples on proteomics-based clinical assays and some clashes with the General Data Protection Regulation (GDPR) in Denmark. Some more examples in other countries would have been much appreciated.

Is the rationale for the Open Letter provided in sufficient detail? (Please consider whether existing challenges in the field are outlined clearly and whether the purpose of the letter is explained)
Yes

Does the article adequately reference differing views and opinions?
Yes

Are all factual statements correct, and are statements and arguments made adequately supported by citations?
Yes

Is the Open Letter written in accessible language? (Please consider whether all subject-specific terms, concepts and abbreviations are explained)
Yes

Where applicable, are recommendations and next steps explained clearly for others to follow? (Please consider whether others in the research community would be able to implement guidelines or recommendations and/or constructively engage in the debate)
Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Clinical proteomics

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Author Response 24 Oct 2023
Filip Mundt

We wish to express our gratitude to the reviewers for their thoughtful and constructive feedback, which has improved the quality of this manuscript. We welcome their input and in the sections that follow, we aim to address each of their comments and suggestions, reflected in the updated manuscript.

1. This Open Letter introduces to the reader some of the challenges we are facing in clinical proteomics. Although most of them are mentioned, the authors mainly focused on very important ethical considerations. I should say that I have missed some more discussion and future perspectives on some other crucial points such as how we could overcome the lack of external validation cohorts and exchange samples, and how to improve the use of more proper comparator groups.

This is indeed a valid point, and in an ideal scenario, every clinical proteomics study should incorporate an external validation cohort. While there have been instances where parallel efforts have corroborated ongoing clinical proteomics studies (as exemplified by Archer et al., 2018, and Forget et al., 2018), the prevailing publishing environment can present challenges in this regard. Additionally, the successful exchange of reference samples between proteomic laboratories, as demonstrated by Mertins et al. in 2018, has been an exception rather than the norm. We have added a brief discussion of this topic, to further address and highlight this issue.

2. The authors describe some examples on proteomics-based clinical assays and some clashes with the General Data Protection Regulation (GDPR) in Denmark. Some more examples in other countries would have been much appreciated.

In our collective experience, we have observed varying degrees of stringency in the application of GDPR regulations to omics data across different EU countries. While we do not possess more specific recent examples, we are concerned that such instances may accumulate unless we, as a community, proactively address the challenges outlined in this work.

Competing Interests: M. M. is an indirect investor in Evosep Biosystems. All other authors declare no competing interests.

Reviewer Report 16 May 2023

https://doi.org/10.21956/openreseurope.17080.r31191

© 2023 Ozlu N et al. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Nurhan Ozlu
[1] Department of Molecular Biology and Genetics, Koc Universitesi, Istanbul, İstanbul, Turkey

[2] Department of Molecular Biology and Genetics, Koc Universitesi, Istanbul, İstanbul, Turkey

Daniel J. Geiszler

[1] Department of Molecular Biology and Genetics, Research Center for Translational Medicine, Koc Universitesi, Istanbul, İstanbul, Turkey

[2] Department of Molecular Biology and Genetics, Research Center for Translational Medicine, Koc Universitesi, Istanbul, İstanbul, Turkey

Mundt *et al.,* provide important insight into the role and possible implementations of proteomic tools in clinics and the ethical aspects of clinical proteomics. The manuscript was written based on a discussion of EPIC-XS members at the Core Technologies for Life Sciences (CTLS) conference.

The authors provide an extensive discussion about the ethical implications of generating and analyzing large-scale proteomic data in clinics. As pioneers in the field, with multiple studies where they successfully apply proteomics in clinics, the authors provide a timely discussion about how to put proteomics on the right track to be used routinely in clinics with the appropriate ethical policies. The authors pointed out some important concepts as re-identification of individuals in a proteomic data set, where one would think that given the dynamic nature of proteins that may not be an issue however, authors provide important evidence from different studies about this possibility. Development in proteomic and computational technologies will likely improve its capacity in the future.

The authors also emphasize the ethical problems surrounding informing the patients about incidental findings which may be useful for the patient's treatment. They raised many points that are shared with genomics studies while also highlighting some unique issues in proteomics.

They also mention an initiation of "controlled access" option for PRIDE for clinical data and its possible impact in the future. This will be interest of several proteomic and clinical researchers.

The authors also make proposals for future steps that can be taken to address clinical and ethical issues in proteomics. These include setting up EU-wide SOPs for MS-based clinical proteomics; developing tailored bioinformatics infrastructure for sensitive data; improving communication between clinicians, patient interest groups, and data access committees; and developing clear guidelines for incidental findings and sensitive data.

We strongly recommend the indexing of this manuscript.

We suggest the following improvements for the manuscript where it is possible before the indexing:

1. The authors talk about the limitations of proteomic applications in clinics but also point out the potential of clinical proteomics when it is combined with machine learning approaches and the power of artificial intelligence. The authors may elaborate on this further and provide by providing some references. More specifically following sentences require references:

   "In our experience, MS-based proteomics can do this in hours and, unlike visual detection, is compatible with machine learning approaches, potentially reducing (human) error. In addition, proteomics may directly provide answers that are normally resolved through subsequent genetic

testing, for example for hemoglobinopathies"

2. The following paragraph is an introductory to the proteomics which needs to be discussed earlier in the text, before going to specific application and should be given with references:

   "*Proteomics additionally enables systematic characterization of the extracellular matrix environment[19], which is difficult with other omics technologies. In many of these advancements, the enrichment and quantification of post-translational modifications are key, which is again a unique area of study for proteomics."*

3. This reference should be inserted properly, *"[28118949]"*

4. Authors mention the importance of international effort in establishing appropriate platforms for clinical proteomics by developing clear guidelines for ethical issues. Perhaps the importance of multi-national effort should also be emphasized in the following sentence:

   "*This needs to be done in collaboration between clinicians, patient interest groups, bioethicists, data protection specialists, and proteomics experts."*

   The countries which do not have much proteomic or computational expertise but rather provide biological samples of some rare or endemic diseases in international collaborations will greatly benefit from ethical guidelines for clinical proteomics that take into account cultural sensitivities.

5. *"Clinics routinely use protein-related information for diagnostic and therapeutic decisions as more than 95% of FDA-approved drugs currently target proteins[2] and hundreds of FDA-approved protein-based clinical biomarkers are readily detectable in blood alone[3]."* - This sentence implies causality ("Clinics routine use protein-related information ... AS ... drugs currently target proteins"), but biomarkers do not need to be drug targets to be useful. It should be slightly rephrased.

6. *"Core Technologies for Life Conference (CTLS; https://ctls-org.eu/) conference in September 2021"* - Should this read "Core Technologies for Life Sciences conference"?

7. "*Several published studies report potential biomarkers of multiple diseases, yet very few, if any, reach clinical application (although proteomics is not alone in this)[4]."* - It would be more accurate here to say "Many published studies".

8. *"ideally taken by longitudinal sampling."* - Is this true in all cases? Why is it true? A reference here might be clarifying for readers.

9. References 5-10 would be more useful if distributed among the individual items they describe in the sentence, e.g., "statistical tools that help to evaluate variance and prove power calculations(5,6)".

10. *"In our experience, MS-based proteomics can do this in hours and, unlike visual detection, is compatible with machine learning approaches, potentially reducing (human) error."* - This

sentence is not clear. What is “visual detection”? Machine-learning based image classification approaches for diagnosis are a large and active area of research, e.g., Ker et al., (2019)[1] .

11. The authors mention experience at two major Danish hospitals (Rigshospitalet and Bispebjerg Hospital) but only provide an experience at one hospital, and it is not clear which of the two hospitals is being discussed.

12. “For instance, a simple blood sample can be used to identify patients at high risk of the tick-borne disease borreliosis, as opposed to risky and invasive spinal tap procedure.” - Can a citation be provided?

13. “Presently, sequence information from proteomics data is less likely to re-identify an individual than genomic or transcriptomic sequence data[28,29].” - This statement is true. However, it is important to note that improvements in computational methods may lead to data becoming re-identifiable (or identifiable) in retrospective studies even if it was not originally re-identifiable (or identifiable) when it was made public.

14. “By contrast, re-identifiable information is information that enables one to match one sample to another in a dataset.” - This sentence comes after a paragraph devoted to re-identifying patients based on proteomics data. It would be better if it were placed before the discussion of patient re-identification.

References

1. Ker J, Bai Y, Lee HY, Rao J, et al.: Automated brain histology classification using machine learning.*J Clin Neurosci*. 2019; **66**: 239-245 PubMed Abstract | Publisher Full Text

Is the rationale for the Open Letter provided in sufficient detail? (Please consider whether existing challenges in the field are outlined clearly and whether the purpose of the letter is explained)

Yes

Does the article adequately reference differing views and opinions?

Yes

Are all factual statements correct, and are statements and arguments made adequately supported by citations?

Partly

Is the Open Letter written in accessible language? (Please consider whether all subject-specific terms, concepts and abbreviations are explained)

Yes

Where applicable, are recommendations and next steps explained clearly for others to follow? (Please consider whether others in the research community would be able to implement guidelines or recommendations and/or constructively engage in the debate)

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Proteomics, computational proteomics, cell biology

We confirm that we have read this submission and believe that we have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Author Response 24 Oct 2023

Filip Mundt

We wish to express our gratitude to the reviewers for their thoughtful and constructive feedback, which has improved the quality of this manuscript. We welcome their input and in the sections that follow, we aim to address each of their comments and suggestions, reflected in the updated manuscript.

We suggest the following improvements for the manuscript where it is possible before the indexing:

1. The authors talk about the limitations of proteomic applications in clinics but also point out the potential of clinical proteomics when it is combined with machine learning approaches and the power of artificial intelligence. The authors may elaborate on this further and provide by providing some references. More specifically following sentences require references: "In our experience, MS-based proteomics can do this in hours and, unlike visual detection, is compatible with machine learning approaches, potentially reducing (human) error. In addition, proteomics may directly provide answers that are normally resolved through subsequent genetic testing, for example for hemoglobinopathies"

The above statement is grounded in our extensive clinical experience with a mass spectrometer for proteomics within this particular clinical context. We have not published on these findings, and are currently unable to provide a reference.

2. The following paragraph is an introductory to the proteomics which needs to be discussed earlier in the text, before going to specific application and should be given with references: *"Proteomics additionally enables systematic characterization of the extracellular matrix environment19, which is difficult with other omics technologies. In many of these advancements, the enrichment and quantification of post-translational modifications are key, which is again a unique area of study for proteomics."*

In response to your query, we have relocated and elaborated on this section, placing it at an earlier point in the document.

3. This reference should be inserted properly, "[28118949]"

The reference has been formatted properly.

4. Authors mention the importance of international effort in establishing appropriate platforms for clinical proteomics by developing clear guidelines for ethical issues. Perhaps the importance of multi-national effort should also be emphasized in the following sentence: "This needs to be done in collaboration between clinicians, patient interest groups, bioethicists, data protection specialists, and proteomics experts."

Ethical guidelines for the management of proteomics data will play an increasingly crucial role, and it is imperative to consider cultural sensitivities when adapting these guidelines to various countries and legislations. As outlined, there is no need to start from scratch, as substantial groundwork has already been laid in the context of DNA/RNA sequencing data. We have added on this important issue of the need of multi-national efforts.

5. "Clinics routinely use protein-related information for diagnostic and therapeutic decisions as more than 95% of FDA-approved drugs currently target proteins2 and hundreds of FDA-approved protein-based clinical biomarkers are readily detectable in blood alone3." - This sentence implies causality ("Clinics routine use protein-related information … AS … drugs currently target proteins"), but biomarkers do not need to be drug targets to be useful. It should be slightly rephrased.

We agree with this comment, and the section has been rephrased to address this.

6. "Core Technologies for Life Conference (CTLS; https://ctls-org.eu/) conference in September 2021" - Should this read "Core Technologies for Life Sciences conference"?

This has been amended.

7. "Several published studies report potential biomarkers of multiple diseases, yet very few, if any, reach clinical application (although proteomics is not alone in this)4." - It would be more accurate here to say "Many published studies".

We fully agree, and it has been changed.

8. "ideally taken by longitudinal sampling." - Is this true in all cases? Why is it true? A reference here might be clarifying for readers.

We argue that since diseases exhibit adaptability, progression, and dynamism over time, then capturing disease dynamics would often benefit from longitudinal profiling rather than relying on single-time-point measurements. Longitudinal studies track changes over time, offering a more accurate understanding of disease progression and response to treatments.

9. References 5-10 would be more useful if distributed among the individual items they describe in the sentence, e.g., "statistical tools that help to evaluate variance and prove power calculations(5,6)".

Good point, and we have now separated those references.

10. “In our experience, MS-based proteomics can do this in hours and, unlike visual detection, is compatible with machine learning approaches, potentially reducing (human) error.” - This sentence is not clear. What is “visual detection”? Machine-learning based image classification approaches for diagnosis are a large and active area of research, e.g., Ker et al., (2019)1 .

Still in clinics today, visual evaluation of test results is a reality. Computerised image analyses are obviously moving forward, but have had some issues, and implementation has not been ubiquitous, yet. We have clarified, and added references.

11. The authors mention experience at two major Danish hospitals (Rigshospitalet and Bispebjerg Hospital) but only provide an experience at one hospital, and it is not clear which of the two hospitals is being discussed.

The proteomics laboratory initiated at the Rigshospitalet, has now moved to Bispebjerg. This has been clarified in the manuscript.

12. “For instance, a simple blood sample can be used to identify patients at high risk of the tick-borne disease borreliosis, as opposed to risky and invasive spinal tap procedure.” - Can a citation be provided?

Unfortunately, we cannot provide a reference currently. This is based on ample clinical experience and yet unpublished data from our groups.

13. “Presently, sequence information from proteomics data is less likely to re-identify an individual than genomic or transcriptomic sequence data28,29.” - This statement is true. However, it is important to note that improvements in computational methods may lead to data becoming re-identifiable (or identifiable) in retrospective studies even if it was not originally re-identifiable (or identifiable) when it was made public.

Indeed, your point is valid. A paragraph has been included to further spotlight this perspective.

14. “By contrast, re-identifiable information is information that enables one to match one sample to another in a dataset.” - This sentence comes after a paragraph devoted to re-identifying patients based on proteomics data. It would be better if it were placed before the discussion of patient re-identification.

Good point. We have moved this section now.

Competing Interests: M. M. is an indirect investor in Evosep Biosystems. All other authors declare no competing interests.