

EDITORIAL



Challenges for omics technologies in the implementation of personalized medicine

Edouard C. Nice

Department of Biochemistry and Molecular Biology, Monash University, Melbourne, Victoria, Australia

ARTICLE HISTORY Received 8 June 2018; Accepted 24 July 2018

KEYWORDS The omics pipeline; proteogenomics; proteomics; precision/personalized medicine; P4 (predictive, preventive, personalized, participatory) medicine

We are currently witnessing the emergence of a new paradigm in health care where the current 'one size fits all' approach, in which diseases are treated based on an average drug response, which does not take into account disease heterogeneity in the larger population, is being replaced with personalized/precision/P4 (predictive, preventive, personalized, and participatory) medicine which utilizes a systems biology approach to accurately diagnose an individual patient's disease at the molecular level, and then use that information to develop individualized targeted treatments specifically for that patient (the right drug at the right dose for the right patient).

Unfortunately, there is currently a common public (and often governmental) misconception that data obtained from genomics alone provides all the necessary information to understand the biology of human health and disease and support precision/personalized/P4 medicine (PM). Clearly this is not the case. Genomics only informs on a person's genetic composition (the blueprint), but not which genes are actually expressed (the parts list) (transcriptome) or functional (the end product) (proteome). While human pathologies are encoded by both our genome and our environment, they are all produced by measurable changes in the human proteome. The genome, encoding approximately 23,000 genes, is essentially static and only changes when a mutation, gene methylation, or translocation occurs. By contrast, the proteome is far more dynamic, more complex, and far larger (estimates suggest up to 1 million distinct proteoforms). This is because proteins can undergo alternative splicing, harbor single amino acid polymorphisms arising from non-synonymous single-nucleotide polymorphisms, or undergo posttranslational modifications.

Following the initial release of the first drafts of the human genome in both *Nature* and *Science* in 2001, it was immediately realized that an in-depth understanding of the human proteome would be required to fully understand the complex biology relating to health and disease, and the *Financial Times* immediately ran a feature article entitled 'The Next Holy Grail, Deciphering the Whole Protein Set,' highlighting the importance of proteomics [1]. Importantly, it must be remembered that almost all current drugs target proteins.

Recent proteomic advances mean that the technologies required to achieve virtually full proteome coverage have now matured. At the mass spectrometry level, comprehensive, reproducible datasets can now be generated rapidly using

sensitive, quantitative, massively parallel, targeted proteomic approaches. In particular this has enabled comparative studies allowing the identification of potential biomarkers or interactome studies revealing key disease-related signaling pathways leading to potential new drug targets. Thus, a recent publication showed single-shot proteomics could provide evidence for more than 90% of the proteome of a human cancer cell line with >6200 proteins present in 10/10 replicates. In mouse brain tissue, >10,000 proteins were detected in only 100 min, with sensitivity extending into the low-attomolar range [2]. This method (BoxCar) also showed a greater dynamic range in human plasma samples, a key hurdle when using this biological matrix. In another example, the top-down proteomic technologies, in which intact proteins are analyzed, were used to detect and quantify mutation-specific consequences of KRAS biochemistry including post translational modifications in human colorectal cancer patients with the same genotype [3]. At the clinical level, proteomics-based techniques are becoming established. For example, the FDA has approved MALDI-TOF-based proteomics instrumentation and assays for microbiological testing, with more than 2000 instruments now in clinics worldwide, as well as the OVA 1 ovarian cancer assay that was based on a SELDI-derived proteomic biomarker panel.

Advances in micropurification techniques, development of fully automated high-throughput array technologies, the availability of highly specific validated monoclonal and polyclonal antibodies [4], the development of technologies capable of preparing highly homogeneous subcellular or tissue preparations (or even single cell analysis (e.g. CyTOF [5])), and advanced bioinformatics further complement the proteomics toolbox. However, even here further improvements can be made as, for example, in the optimization and validation of improved protein extraction protocols allowing even deeper mining of the proteomes [6]. These technologies, in turn, will form part of the omics pipeline comprising techniques such as genomics, epigenomics, transcriptomics, proteomics, peptidomics, interactomics, metabolomics, and microbiomics that together will support PM. There is no doubt such an approach to PM is both viable and effective as evidenced by several recent studies. A comprehensive personal omics characterization (Integrative Personal Omics Profile), which combined genomic, transcriptomic, proteomic, metabolomic, and

autoantibody profiles of a single individual over a 14-month period, revealed dynamic molecular and medical phenotypes including a risk for type 2 diabetes [7]. In an extension of this study (hPOP trial), volunteers at the annual international Human Proteome Organization (HUPO) meetings, which alternate between the United States, Europe and Australasia, supply blood, urine, and fecal samples at the meeting itself, following strict standard operating procedures, for a comprehensive analysis involving whole genome sequencing, transcriptomics, proteomics (including analysis of the microbiome), and metabolomics, providing the potential for longitudinal (regular attendees) and ethnicity studies. In another example, the Institute of Systems Biology launched the 100K Wellness Project in 2014 [8], with the goal of making detailed genomic, proteomic, epigenetic, metabolomic, and phenotypic measurements on 100,000 apparently healthy volunteers several times a year, examining blood, saliva, and stool as well as other physiological and psychological parameters to study the initiation and progression of all common diseases.

So we have a functional toolbox: where are the roadblocks? Unfortunately, they are multifarious [9]. Fundamental questions are those of clinical utility, diagnostic accuracy, and evidence-based pathology to support proteomic diagnostics. To help address this, in the first major change in operational structure since HUPO was launched in 2001, HUPO has recently announced that Pathology will join Mass Spectrometry, Antibodies, and Knowledge Base as the fourth resource pillar of the organization. This acknowledges the role pathology will play in the delivery of proteomics driven biomarker assays of clinical utility. It is proposed that the new Pathology Pillar will coordinate the identification of key unmet needs in clinical medicine, develop guidelines and standards for 'fit for purpose' validated clinical assays, promote awareness of best practice, and coordinate access to appropriate state of the art biobanks (which are a critical but rapidly depleted resource) and associated data.

A second fundamental question is that of credibility. Emerging technologies are often burdened with over enthusiasm or 'hype,' and that has unfortunately been the case with many of the omics technologies. In the early days, proteomics was thought by many to have overpromised and under-delivered [10], especially in the area of biomarker discovery. Similar hype concerns are now being voiced about liquid biopsies and next-generation sequencing (NGS), especially as the technical platforms are not yet stable, laboratory quality assurance programs are still in their infancy, currently there are no gold standard methods and protocols for their evaluation, and data interpretation capabilities are still very underdeveloped [11].

Data analysis and storage is another major roadblock, with the size and multiple different formats of datasets generated posing a considerable technical and statistical challenge. There are no universal protocols for modeling, comparing, or benchmarking the performance of the various data analysis strategies that are often a 'black box' for the user. Thus, ideally, potential new biomarkers should stand the test of multiomics data triangulation before they can be prioritized for clinical applications [12]. It is perhaps not surprising that some caution has been advised for the analysis of the Big Data generated [13,14]. For example, it has been suggested care must be

taken when using techniques like Artificial Intelligence (AI). A perceived problem here is the availability and accuracy of the data used to help train these systems. Developing robust AI engines requires large datasets, but this may not always be the case for health care. Data storage may also be a problem. It has been estimated that countries with large populations and emerging economies such as India and China and India will be generating zettabytes (10^{21}) to yottabytes (10^{24}) of health-related data each year [15]. Although cloud-based systems offer a potential storage solution, concerns have been expressed around security and privacy [16].

What about problems in the practical implementation of personalized medicine? A major question here is sustainable funding. To deliver the required whole systems approach, large international multidisciplinary teams will often be required (we are already seeing this trend in many of the recent publications emanating from key proteogenomic laboratories). A number of major international initiatives around proteomics and PM have been initiated, illustrating the importance of taking a multi-technology approach. In 2015, President Obama announced a precision medicine initiative in his State-of-the-Union address, with funding of more than US\$200 million, which now includes proteomics as a key component. Building on this, the Cancer Moonshot effort was launched to use proteogenomics to specifically accelerate cancer research to make more therapies available to more patients, while also improving early stage detection where treatment is far more successful or even curative. As part of this, the International Cancer Proteome Consortium (ICPC) has been formed, with coordination through HUPO. Australia was the first country to formally join the ICPC and a further 11 other international countries have now signed MOUs (Canada, Sweden, South Korea, the UK, India, Japan, Switzerland, Taiwan, the United States, Germany, and China). China clearly intends to play a major role in defining the human proteome and has recently released a 5-year plan which commits more than \$US9 billion to fund precision medicine research. It is essential that there is sustainable long-term funding. However, many funding bodies are reluctant to fund outside their own borders, and a new funding paradigm to address this is urgently required [17]. It is essential to ensure that funds are available not only for large-scale cataloging exercises (such as the HUPO CHPP project), but also for the multinational multidisciplinary research-based projects mentioned earlier and, importantly, the 'Little Science' projects often addressed by specialist laboratories [18]. However, it has been suggested that taxonomies for omics studies are still limited, requiring a re-think in the way in which various omics science studies are classified, prioritized, and ranked [19]. Reflecting funding availability, many of the essential technology platforms may not be readily available in third world countries, which could lead to a two-tier medical system: improved health benefits should flow to all.

To maximize potential returns scientifically and economically, it is essential we ensure that the potential health and financial benefits of proteomics are effectively and efficiently incorporated into our long-term planning strategies. Not only will the optimized health care that arises from PM bring with it large savings to health budgets, but financial benefit will also accrue from the associated IP (which will include new

biomarkers, drugs, and drug targets), new employment opportunities which will arise from the provision of new clinical services required to support PM, and new manufacturing opportunities in areas like GMP drug manufacture, microfabrication, and instrument development.

To ensure that the role of proteomics and other omics technologies is both understood and effectively incorporated into a workable international health policy, a global working party on PM should be established as a matter of urgency. This should be composed of an international multidisciplinary team of informed experts including scientists, clinicians, educationalists, politicians, professional organizations and industry to coordinate research activities, to optimize efficient roll out and support of required technologies, to ensure that appropriate infrastructure and sustainable long-term funding is available and that proteomics is appreciated as a core component of PM.

Acknowledgments

I gratefully acknowledge my friend and collaborator, Prof Mark Baker (Macquarie University, Australia), co-chair of the HUPO HPP, with whom I have spent many enjoyable hours discussing how best to advance the “Omics” technologies and precision medicine.

Declaration of interest

EC Nice is the co-chair of both the Human Proteome Organization (HUPO) Cancer BD and the HUPO Pathology Pillar. The author has no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

Reviewer disclosures

Peer reviewers on this manuscript have no relevant financial or other relationships to disclose.

Funding

This manuscript has not been funded.

References

Papers of special note have been highlighted as either of interest (*) or of considerable interest (***) to readers.

- Omenn GS. Plasma proteomics, The Human Proteome Project, and cancer-associated alternative splice variant proteins. *Biochim Biophys Acta*. 2014;1844:866–873.
 - Meier F, Geyer PE, Virreira Winter S, et al. BoxCar acquisition method enables single-shot proteomics at depth of 10,000 proteins in 100 minutes. *Nat Methods*. 2018 May 7.
 - Ntai I, Fornelli L, DeHart CJ, et al. Precise characterization of KRAS4b proteoforms in human colorectal cells and tumor reveals mutation/modification cross-talk. *Proc Natl Acad Sci U S A*. 2018;115:4140–4145.
 - Uhlen M, Bandrowski A, Carr S, et al. A proposal for validation of antibodies. *Nat Methods*. 2016;13:823–827.
 - Nowicka M, Krieg C, Weber LM, et al. CyTOF workflow: differential discovery in high-throughput high-dimensional cytometry datasets. *F1000Res*. 2017;6:748.
 - Karthikaichamy A, Deore P, Rai V, et al. Time for multiple extraction methods in proteomics? A comparison of three protein extraction methods in the eustigmatophyte alga *Microchloropsis gaditana* CCMP526. *OMICS*. 2017;21:678–683.
 - Chen R, Mias GI, Li-Pook-Than J, et al. Personal omics profiling reveals dynamic molecular and medical phenotypes. *Cell*. 2012;148:1293–1307.
 - Institute for Systems Biology: 100K Wellness Project. Available from: <https://systemsbiology.org/research/100k-wellness-project>
 - Nice EC. From proteomics to personalized medicine: the road ahead. *Expert Rev Proteomics*. 2016;13(4):341–343.
 - Huber LA. Is proteomics heading in the wrong direction? *Nat Rev Mol Cell Biol*. 2003;4:74–80.
 - Wilson BJ, Miller FA, Rousseau F. Controversy and debate on clinical genomics sequencing-paper 1: genomics is not exceptional: rigorous evaluations are necessary for clinical applications of genomic sequencing. *J Clin Epidemiol*. 2017;92:4–6.
 - Pecak M, Korošec P, Kunej T. Multiomics data triangulation for asthma candidate biomarkers and precision medicine. *OMICS*. 2018;22:392–409.
 - Lee CH, Yoon HJ. Medical big data: promise and challenges. *Kidney Res Clin Pract*. 2017;36:3–11.
 - Househ M, Aldosari B. The hazards of data mining in healthcare. *Stud Health Technol Inform*. 2017;238:80–83.
 - Andreu-Perez J, Poon CC, Merrifield RD, et al. Big data for health. *IEEE J Biomed Health Inform*. 2015;19:1193–1208.
 - Esposito C, Castiglione A, Tudorica C-A PF. Security and privacy for cloud-based data management in the health network service chain: a microservice approach. *IEEE Communications*. 2017;9:55.
 - Poste G. Biospecimens, biomarkers, and burgeoning data: the imperative for more rigorous research standards. *Trends Mol Med*. 2012;18:717–722.
 - Alberts B. The end of “small science”? *Science*. 2012;337:1583.
 - Pirih N, Kunej T. Toward a taxonomy for multi-omics science? Terminology development for whole genome study approaches by omics technology and hierarchy. *OMICS*. 2017;21:1–16.
- **This manuscript illustrates the potential for rapid deep mining of the proteome.**
 - **This manuscript describes a top-down proteomics workflow that can detect and quantify mutation-specific posttranslational modifications related to KRAS biochemistry.**
 - **A report from an international working group for antibody validation to formulate the best approaches for ensuring quality and antibody reproducibility.**
 - **A key reference demonstrating the potential of precision/personalized medicine.**
 - **A critical appraisal of the use of NGS technologies as screening, diagnostic, or prognostic tools.**