

EDITORIAL

Exploring the Human Plasma Proteome

The Human Proteome Organization (HUPO) has launched several major initiatives to explore the proteomes of liver, brain, and plasma and to generate informatics standards and large-scale antibody production. This issue of *PROTEOMICS* presents 28 papers from the pilot phase of the Human Plasma Proteome Project. The papers represent a combination of collaborative analyses of HUPO PPP reference specimens and many lab-specific special projects, both experimental and analytical. Plasma (or serum) will be the preferred specimen for non-invasive sampling of normal individuals and patients for biomarkers discovered and validated to reflect physiological, pathological, and pharmacological phenotypes. As is well-recognized, however, plasma presents enormous challenges due to the extreme complexity reflecting potentially all proteins in the body and their isoforms, at least ten orders of magnitude range in concentration, intra-individual and inter-individual variation, and non-standardized methods of sample processing. Furthermore, the inherent limitations of incomplete sampling of peptides by mass spectrometry and potentially high error rates of peptide identifications and protein assignments with various search algorithms and databases lead to low concordance of protein identifications even with repeat analyses of the same sample.

Together these papers examine many features of specimen handling, depletion of abundant proteins, fractionation of intact proteins, fractionation of tryptic digest peptides, and analysis of those peptides with various MS/MS instruments. The subsequent steps of matching spectra to peptide sequences with a variety of search algorithms that have numerous and often unspecified parameters, followed by alignment of peptide sequences with proteins in a variety of protein or gene databases, laden with uncertainties and redundancies, are explored in considerable depth. Especially for longitudinal and collaborative studies, the periodic issuance of remarkably modified versions of the databases creates a moving target for protein identification and annotation. These reports provide a revealing snapshot of the variety of proteomics technologies as of 2004 and the variable depth of analysis across laboratories. Subproteomes based on molecular weight or glycoprotein enrichment yielded additional findings. The extensive annotations show that even our present methods are capable of detecting large numbers of low-abundance proteins of great biological interest from essentially all cellular compartments. As more powerful technologies are applied, we can expect ever more extensive identifications. The high proportion of genes which appear to generate detectable splice isoforms complicates the protein identifications, yet helps to clarify, together with myriad post-translational modifications of proteins, the basis on which humans can have such complex phenotypes with a surprisingly small complement of genes (latest Human Genome Project estimate about 22,000 protein-encoding genes).

The PPP datasets offer a rich resource for many follow-on analyses, as illustrated here, and an anchor for new studies. We provide a Core Dataset of 3020 proteins identified with two or more peptides and remaining after integration of protein matches which cannot be distinguished with the available peptides. Numerous other “cuts” of the data can be made. Extensive primary data and analytical results are



Gilbert S. Omenn

available at both the University of Michigan and European Bioinformatics/PRIDE websites. A special feature is the beginning of independent analyses from the raw spectra or peaklists from multiple labs. These independent analyses eliminate the high variability of different search algorithms, different databases, and different investigators, though each independent analysis has its own peculiar attributes. We also provide some comparisons with published datasets, including the extent of overlap in protein identifications. Meta-analysis of separate studies has similar challenges to those experienced here in the integration of datasets from collaborating laboratories in the PPP.

The PPP datasets set the stage for planning the next phases of the PPP. The next phases will be aimed at generating standardized operating procedures, creating an active bioinformatics effort that captures and compares new reports, and stimulating combinations of technologies and informatics with higher throughput, greater dynamic range, and lower error rates. Given that different mass spectrometers utilize different search algorithms, we will need to establish equivalent search algorithm parameters and filters; some progress has been made in the pilot phase. Replication of results deserves a special priority, since it is very awkward to seek disease-specific differences between specimens from different individuals when the same specimen in the same lab typically has not more than 40 to 50 percent concordance on repeat analysis. On the path to biomarker discovery and validation for diseases of the liver, brain, and other organs, experimental analyses of plasma samples from the same people and animals whose organs are studied will be essential.

Detection of isoforms and quantitative measurement of proteins also warrant high priority, given the importance of post-translational modifications of many kinds. The PPP has demonstrated through the use of commercially-available quantitative immunoassays that protein abundance is the prime factor in the probability of detecting peptides of a given protein, modified, of course, by various protein-specific features and experimental protocols.

The spirit of collaboration in the PPP has been splendid, and the substantial commitment of so many investigators and sponsors to this pilot phase of the HUPO Plasma Proteome Project has been admirable. As a work-in-progress, the PPP has generated productive discussions at many HUPO, U.S. HUPO, and other scientific meetings. On behalf of our Executive Committee and Technical Committees, I thank everyone involved. We welcome others to express interest and influence our future work, starting with workshops and presentations at the 4th HUPO World Congress in Munich.



Gilbert S. Omenn
University of Michigan
Ann Arbor