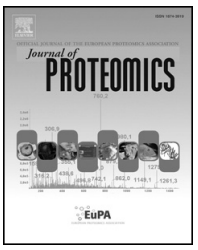
Available online at www.sciencedirect.com

ScienceDirect

www.elsevier.com/locate/jprot



Editorial

Can proteomics fill the gap between genomics and phenotypes?

The Human Genome Project, completed a decade ago, deciphered the human genome with the current number of human genes being 20,135. The successor of the Human Genome Project, the Human Proteome Project, was launched in 2010 at the Annual HUPO Meeting in Sydney, Australia. The HUPO international consortium was formed to identify and understand the function of all proteins in the human body and to map the human proteome on a disease- and chromosome-centric basis. By knowing all the proteins and their interactions and roles in pathways, much will be learned about the fundamental ways that cells and tissues function. Using the techniques of proteomics, modificomics and bioinformatics, the proteins coded by each of the genes on each chromosome will be methodically identified and their modifications and role in cells and tissues deciphered. The proteome in each major human disease will likewise be analyzed. Each country in this consortium will classify and characterize the proteins associated with one or more of these identified diseases and chromosomes. Canadian investigator teams have identified chromosomes 6 and 21 and a goal of the 2013 Canadian National Proteomics Annual Meeting was to identify a disease for detailed analysis.

Completing the Human Genome Project led to more questions. For me one of the big questions is: Why are there so few genes in man? A cabbage for instance has double that of humans, with 41,174 genes and an articulated London double decker bus has over 24,000 different parts, yet performs only a limited set of functions, and as mechanics will readily regale, it most certainly cannot self-repair. Thus, a critical question is that with only 20,135 genes and following the one gene one protein assumption, how sufficiently different need these links among proteins and functions be to generate the human life form? Many other organisms have similar genes and similar numbers of genes as humans. E.g. *Drosophila* has 15,016 genes and *Caenorhabditis elegans* has 20,470 so where does the extra information come from that makes humans distinct from other organisms?

The answer to this big question lies in what has become one of the hottest areas of proteomics in the past few years, 'modificomics': the study of how different forms of the same protein are produced by splicing and how proteins are modified after synthesis by post-translational modifications (PTMs). These modifications, for example, change the chemistry of the amino acids of a protein, e.g., acetylation, phosphorylation, methylation, glycosylation and nitration and can minimally truncate proteins or entire sections of proteins are removed. These modifications then can change the function of each protein, for example, changing a "go" signal to a "no-go" signal in a protein, moving proteins from one part of the cell to another where they may play different roles, or altering the half life of a protein. Thus, the number of proteins can be considered to increase 5-fold by considering all splice forms per protein. When the ~300 PTMs are considered the number of distinct protein chains rises to around greater than 500,000. On top of this precise proteolytic processing can convert the parent chain to two or more stable daughter chains giving rise to up to potentially 1,000,000 protein chains. By considering differences in the generation of these protein numbers in different species then it is less of a step in understanding where the extra information comes to differentiate a fly from a worm from a human. However, understanding the genetic sequence alone does not provide this information. We need to understand the protein. But how are these modifications coded? Understanding this will go a long way to filling the lack of understanding of the linkage between the genome and proteome; between the cells and whole organisms; and essential differences in health and disease. When these changes are misregulated diseases such as arthritis or cancer may result.

The incidence of cancer and chronic inflammatory diseases is growing with an aging population and this will result in significant costs to society and the health care system. In recent years, the linkage of inflammation to cancer is being increasingly recognized as a fundamental factor in carcinogenesis and metastasis. Early detection of chronic inflammatory diseases and cancer should lead to timely and effective treatment options and thus, greatly improves clinical prognosis with decreasing morbidity and mortality. The current

1874-3919/$ – see front matter
http://dx.doi.org/10.1016/j.jprot.2014.02.025

lack of precise methods of diagnosis, including biomarkers, impedes early and accurate clinical diagnosis. This lack of timely and accurate diagnosis delays appropriate and early treatment and thus results in patient morbidity, loss to the economy and, in cancer, death through metastases. Hence, we must develop and introduce cost-effective, minimally invasive, new diagnostic methods including biomarker peptides and proteins.

Recent advances in genomic and proteomic analyses have greatly increased the number of genes and proteins identified as potential biomarkers of disease. However, in order to translate candidate disease proteins to the clinic as biomarkers, we need to understand both their function and also differential protein expression and the more than 300 post-translational modifications of human proteins. These modifications of key regulatory molecules often increase disease activity or can protect against disease. However, the modification of signal pathway components by precise post-translational modification of a protein turns bioactivity on or off in ways that are unpredictable from transcript or conventional proteomics analyses alone.

Currently, by not considering PTMs this may lead to the wrong diagnosis and hence wrong treatment, poor patient outcome and unnecessarily increasing health care costs. If these modified disease proteins can be detected in a timely manner, this will aid patient diagnosis as specific disease-derived biomarkers with the added benefit of reflecting the actual disease state. However, after years of research and millions of research dollars there are only a few biomarkers in clinical use. New approaches are therefore needed which was a focus of the symposium. Hence, we propose that PTM modified peptides can form the basis for high value mechanistically informative biomarkers of disease. Robust biomarkers will revolutionize early disease detection, thus reduce health care costs, save lives and improve the health and well-being of populations and in addition, stimulating the biomedical industry and generating employment in highly sought after biotechnology and medical diagnostic companies.

An unexpected benefit in enriching for post-translational modified peptides from proteins is that such modified peptides will have different and sometimes more favorable ionization and fragmentation properties to the parent tryptic peptide. With 3844 proteins of the 20,135 human proteins designated as missing, meaning they are lacking experimental evidence for their existence at the protein level, enriching and characterizing such modified peptides can be used to provide evidence for a missing protein. Indeed, this strategy was presented at this symposium where six "missing" proteins were identified in human erythrocytes from their mature N termini or endoprotease cleaved neo-N-terminal peptides [1]. Protease cleavage of proteins generates semitryptic peptides and in some of these cases cleavage removes peptide segments that render the peptide too long or too negatively charged for routine mass spectrometric identification. This was exploited to identify missing proteins and it can be envisioned that enriching for other modifications will render a different set of peptides analyzed from a protein than is commonly analyzed.

To address these important questions, the Canadian National Proteomics Network hosted a symposium entitled "Can Proteomics Fill the Gap Between Genomics and Phenotypes?" on April, 20–24, 2013 in Vancouver, BC. Leading Canadian and International scientists and trainees attended. In addition, the meeting was linked with the inaugural meeting of the Pan-American Human Proteome Organization. The results and insights from this symposium are being made accessible to the general scientific community by this special dedicated issue of the Journal of Proteomics.

REFERENCE

[1] Lange P, Huesgen PF, Nguyen K, Overall CM. Annotating N termini for the Human Proteome Project: N termini and Nα-acetylation status differentiate stable cleaved protein species from degradation remnants in the human erythrocyte proteome. J Proteome Res 2014. http://dx.doi.org/10.1021/pr401191w [in press].

Christopher M. Overall
Centre for Blood Research, University of British Columbia, Vancouver, BC Canada
Corresponding author at: University of British Columbia, Centre for Blood Research, 4.401 Life Sciences Institute, 2350 Health Sciences Mall, Vancouver, British Columbia V6T 1Z3, Canada.
Tel.: 1 604 822 2958; fax: 1 604 822 7742.
E-mail address: chris.overall@ubc.ca; http://www.clip.ubc.ca