BSPR Panel Discussion
1800-1845, 13 July 2011, Hinxton Hall Conference Centre
The Wellcome Trust Genome Campus, Cambridge

The next steps in proteomics – What is needed and what can be achieved?

Chair: David O'Connor
Panel: Sam Hanash, Gil Omenn, Michael Taussig

1. Human Proteome Project (HPP) question

Concern it could eat up funds for other projects.

Gil Omenn – there are 3 pillars in the HPP: mass spectrometry; antibodies; and integrated knowledge-base.

Swiss-prot has data on 13000 proteins. We have a lot of info about the proteins coded for by genes. But we don't have an adequate start on post translational modifications, splice variants, so this is the backdrop for a more ambitious project.

Lessons from the Human Genome Project (HGP) for HPP. At the start of the HGP there was a small science - big science debate: concern about the impact of the HGP on small science. In 1980s, ridicule against human genome project was dripping, but it was proven wrong. The human genome project has generated a range of findings, tools, stimulated underlying technology. Similarly in proteomics, there have been technology surges in mass spectrometry and reagents.

We have already had thrusts in biology and disease-based initiatives. Organ-based and tissue-based initiatives feed into chromosome agenda.

Human centric? When initiatives to do research on model organisms were proposed to HUPO, HUPO said yes immediately (http://www.hupo.org/research/imop/). The community is very welcoming to such initiatives and enthusiastic. Similarly with the HGP, in the early years people said, 'What about all those other species?' Research on comparative genomics has added value – the unity and diversity of species is reflected in their genomes.

But that said, strategically a largely biomedical base will be more attractive to foundations, and the government agencies in various countries which have a focus on humans.

Mike Taussig: I only heard about HUPO's initiative on model organism proteomes (iMOP) today. The HPP should be more than a cataloguing exercise and have biological point. The Human Protein Atlas is part of this. Prioritization of how you go about it is very important. The logic of why a particular approach should be adopted needs to be explained.

Could be chromosome-, gene-, disease-, or tissue-centric? Not sure why chromosome-centric has more logic than the others.

Need to be careful not to duplicate existing projects.

Also organisation of this internationally will be a challenge – how to organize – all HUPO? How much will it cost? Through national funding agencies? How many will take part in this?

Sam Hanash – HPP has already started. 100s of mass spectrometers are cranking out the data – thousands of labs. The compelling challenge is to figure out how to capture all the data and make it available, so that investigators have at their fingertips all the data that is available. If we can capture that, it would be at a minimum cost but maximum return on investment.

At the Fred Hutchinson Cancer Research Center we are handling data capture like this. We have organized a database to make sense of our data. There are other 'omics. We have to put proteomics data in context with these other data. We have constructed a database on cancer – has all proteomics data and other 'omics data. RNA sequences etc. for our own use at present. Tommy Nilsson this morning described a protein I have never heard of before. If I wanted to find out more about this – what do we already know about this? I put this into our database. I looked at our data across cell populations we have analysed and downloaded what we have.

[Sam demonstrates this search.] Turns out we have some data on this protein. We have data across many populations. Biology can be extracted without doing a single experiment. Across a number of cell lines and populations we have studied, there is a lot of variability in how many times we have sampled this protein. We could then look at gene expression for this same tissue type. Now we are looking at RNA to see if the biology from protein-to-RNA makes sense. We see it expressed in same populations. It is most expressed in cell line xxxx. This is the one where we have the most abundance at protein level. Mass spec data is reflecting biological variability. This is what I would like to be able to do – download the data from the hundreds of labs working on proteomics. And the HPP is the most compelling project we could do.

2. Data quality question

David O'Connor– what about data quality? HPP will be much bigger than HGP. Will there be data factories? Or if smaller labs participated, then how to maintain quality?

Gil Omenn – SRM is a recent development. We used to detect 'same old same old' and never get to the others. So a new technology was developed and about 3 years ago Ruedi Aebersold went to EU for substantial, but not extraordinary funds, and from USA sources too, and from this we are on the verge of not just having spectra for every key protein, but also having labelled peptides. It's practical but not outrageously expensive. Changes the ball game. Same with the Protein Atlas. Buying antibodies is killing us. Money is wasted on unreliable reagents. We need to get experience together and make companies justify their products. This benefits every investigator, not just giant factory labs. Comes from one lab, but benefits all.

All of the research should be driven by biology and medical questions. The National Institutes for Health and the Medical Research Council are interested in diseases and responses to therapies and environmental pressures. So the risk that we will only do technical development and not biological questions is not real.

David O'Connor – I can see tech advance and reagents, but there are so many diseases one group cannot do them all, so what is the model you see? Is it a distributed model? And if so, how to maintain data quality?

Gil Omenn – it's not top down. Won't be like Frances Collins (HGP). Data quality is critical. This issue came to fore in USA recently: gene expression studies predicting drug action in participating patients – some papers stimulated many labs to follow their lead, and now all retracted. How does that happen? [some of the papers were written by] 10 authors in single institution. If they don't fully understand what each other does then they shouldn't put their name on the paper. Many fields are like this. We have to earn the trust of funders and of the public. We need transparency, quality, reproducibility and sufficiency of data. We are pushing our reviewers as hard as we can. Universities say it's the responsibility of individual scientists. This is not good enough. Everybody has to step up.

3. Open to floor - Data Quality discussion continued

A: We have demonstrated that as long as we can have the raw data in central repository and analyse it, that something that will work. The quality of the papers submitted is relatively independent of the type of mass spec used.

I think that personally I would prefer to have it as a community project for the very reason that, in contrast to HGP, the HPP is much more open and will be defined as we go into the future. Studying all of the isoforms etc. is the work of generations.

B: I fully agree, but data management and interpretation will be key. Will should really use the impact of this and make the general population realize we can do this

Gil Omenn – some people don't have bioinformaticians. We need training and finding people who can be expert collaborators. ProteomeXchange is going to raise the standard for data and for connecting the pieces. Connection between PRIDE and Peptide Atlas has been growing and is available. It's up to the community to put in adequate metadata. People don't do it and complain about other people's datasets. There are ways of making sure datasets can be saved in multiple formats and reused liked Peptide Atlas. New generations of tools for having access to fully annotated datasets.

C: Providing and collecting all the data needs not only to be well annotated but well curated. We need to have common goal

Gil Omenn – but we have to start by having the data properly labeled, organized and reviewed at the start

4. The future of mass spectrometry?

David O'Connor – moving on. Another question was, 'Does mass spectrometry have a long term future in proteomics?'

Mike Taussig – MS is the gold standard, technically. As long as we can develop enough sensitivity would like to see it combined with other methods e.g. affinity methods, mass spec imaging. A really wonderful thing if it can do the job. Adaptation of mass spec to go into diagnostic field in hospital settings. Once we have gone beyond cataloguing, what else can you do with it?

Gil Omenn – Denis Hochstrasser was very enthusiastic about desktop mass specs. There is a great future for mass spec, but for what? Good for discovery and SRM and MRM data. But it needs too high concentration to get to lower abundance proteins. Needs to go

down at least 2 orders of magnitude for clinical applications. Now we are still hoping nanotechnology - nanoprobes - will go down those low logs. From a discovery point of view MS is good

D : MS in clinical setting. MS can be used for metabolomics studies for delivering biomarkers. How are proteomics performing against these other technologies? What does proteomics have to show in terms of clinically useful biomarkers? We are getting compared with metabolomics and next generation sequencing, and we have very little to show for it.

Sam Hanash – Metabolomics is so much easier to do! To use the mass spec for this! In a comprehensive fashion.

Gil Omenn – I don't want to talk down another field. I am invested in metabolomics myself. But it is very disappointing that a typical metabolomics publication has only 60 metabolites. It's shocking after the promises of several years ago – sounds familiar – but it will come.

Rob Beynon – We do not have a parallel system like in next generation sequencing. We have a serial system - we need massively parallel proteomes. I should like to see an array of protein …

[5. The discussion moves back to HPP]

Rob Beynon – The Human Proteome Project is really exciting. But what about a discussion about the proteome of wheat or rice or mosquito? In terms of the sum total of human suffering these are also very important. The drain on humanity is starvation, and food is going to disappear on this planet.

Gil Omenn – On the chromosome problem. I was not an advocate for a chromosome based project. But the whole human proteome is too great for any single group. And around the world there is enthusiasm for a sub project like a single chromosome. HGP was done in that way. The question from a biological point of view is, 'Is it silly or might it be interesting?' Proteins are coded by genes. Chromo 17 – we have started on a project on this in US, led by Bill Hancock, because it happens to have breast cancer genes on it. And it is a fair question – how many genes are there on this chromosome? Approx 1100 protein coding genes. We already know something about 800-900 of those. Ron Beavis, Mike Snyder are on board. We are going to make a model of what you can do to get launched – lay out a picture of chromosome - genes, SNPs, transcription, proteins – you can see – is there cis regulation?

Koreans chose chromosome 13. Is a way to take a bite of the apple - and maybe find something interesting.

David O'Connor – The Chinese chose chromosome 8 because it's a lucky number for them

Gil Omenn China changed to chromosome 1. The thing is for 3 or 5 groups to show it can be good. There is no need for a rush.

5. Inspiring future proteome scientists?

David O'Connor: how would you inspire the proteome scientists of the future?

Sam Hanash – Provide tools – as I said earlier – how do I get the lab to do the research on the proteins I just discovered? Tools, reagents etc

Mike Taussig – Students have to be excited they are working on the most important problem there is. But an important problem and exciting technology will inspire them. With proteomics we need a good understanding of proteins in terms of structure and not just prevalence. Crystallography could be good

Gil Omenn – Plants and microbes are important. One area of human disease that proteomics has not made as much impact on is infectious diseases. Funding in USA is very high for this – HIV, acute infections. We have learned that our bodies are 90% microbial cells. How does this affect our nutrition and drug responses? These exciting areas are not overcrowded yet.

A: On the question of young investigators and how to inspire them. If they see success they will be interested in following it. Also, the technology needs to be accessible. In Europe imaging is organised around …. How can we focus the know-how collectively …?

Gil Omenn - There is a philosophical issue here. We have the capacity to identify many proteins, but we don't know how to talk about it very well. We talk about the few. This doesn't capture the richness of the system. We have to learn to think about - to visualize - function on a grander scale than one protein at a time.