



8TH SIENA MEETING

FROM GENOME TO PROTEOME:

INTEGRATION AND PROTEOME COMPLETION

Siena, Italy, August 31st- September 4th, 2008

Auditorium Giurisprudenza e Scienze Politiche

REPORT

The 8th Siena meeting “From genome to proteome: integration and proteome completion” took place on Aug. 31st- Sept. 4th, 2008, in the medieval town of Siena, Italy.

About 400 participants from all over the world attended the meeting, which was organized with plenary lectures and poster exhibition in the morning and two parallel sessions in the afternoon. The diversity of the presentations addressed the theme of the meeting, showing the state of art of proteomic approaches and the current efforts to achieve proteome completion. I report here my personal highlights from some plenary lectures.

After the nice Welcome Party held under the tower in Piazza del Campo, the conference started the following day with an opening lecture “A complete set of annotated human proteins in UniProt/SwissProt”, from Amos Bairoch (*Swiss Institute of Bioinformatics, Geneva, Switzerland*). The speaker addressed the main theme of the meeting, “integration and proteome completion”, announcing the completion of the manual annotation of all the known 20,500 human protein-coding genes, according to Swiss-Prot quality standards. The second part of the talk highlighted the limits of our knowledge about the human proteome. In fact, the annotated 20,500 human genes could encode one million different proteins, largely not characterized so far. Thus, it is clear that the current information about protein variants, structure, modifications, sub-cellular localization, etc. still need to be completed and integrated to obtain a comprehensive understanding of protein functions and interactions.

Jean-Michel Claverie (*Mediterranean Institute of Microbiology, Marseille, France*) continued the first plenary session with the intriguing lecture “Mimivirus: endless surprises”. Claverie presented surprising features of Mimivirus, a giant and complex “mimicking microbe” virus, that goes beyond the known limit between viruses and cellular organisms and that, for these reasons, intrigues virologists and evolutionists. Even though the analysis of its proteome led to the identification of proteins significantly similar to known homologous, a relevant part of the proteome did not exhibit a clear database match and is totally uncharacterized. The origin and function of the corresponding ORFs are quite controversial and remain to be explored.

The following day, Werner Arber (*Biozentrum, Basel, Switzerland*), Nobel laureate in Physiology / Medicine in 1978, focused his plenary lecture “Genetically encoded

generators of genetic variants” on the analysis of genetic variation sources. The speaker showed that rare isomeric forms of biological macromolecules (including proteins) may be important factors as occasional source of new biological activity. He thus invited the audience to pay attention to these unstable forms and to avoid to apply a “gene-centric” approach to genome studies.

The third plenary session, entitled “Quantitative and Functional Proteomics”, was of particular interest for me. The first speaker, Sung-Hou Kim (*Berkeley Structural Genomics Center, California, USA*), talked about protein structure universes and genome analysis in the lecture “Classification and evolution of structural and sequence proteomes”. An innovative approach (called feature frequency profile, or FFP) was applied to the comparison of protein structures of different organisms, leading to the observation that only a very small number of fold classes (only 4) have been conserved among the different species. Interestingly, despite this strong architectural constrain, the molecular functions are surprisingly spread all over the 4 fold classes, suggesting that an opportunistic match between protein structures and functions occurred during evolution.

In the second part of the talk, the speaker showed the results obtained by applying the FFP approach to the comparison of whole genome sequences (coding and non-coding regions) or whole genome coding sequences. In both cases, the approach drew the “demography” map of the analyzed organisms, providing a global view about their evolutionary relationship. Interestingly, the fact that these results have been achieved also from the analysis of the whole genome sequences (thus containing up to 99% of non-coding DNA), strongly supports the thesis that the so-called “non-coding” regions of the genome have an unknown function that still need to be elucidated. The speaker concluded the talk remarking the need for a not “gene-centric” model that could be helpful for evolution analyses.

The following speaker, Leigh Anderson (*Plasma Proteome Institute, Washington DC, USA*), addressed the problem of biomarker validation in his interesting talk “Tools for accurate measurement of candidate biomarkers in plasma: development of multiplexed MS-based SISCAPA assays”. Even though accessible biomarkers exist in blood and are able to define all disease states in human population (that is the “strong biomarker hypothesis”), proteomic studies did not successfully contribute to the generation of new diagnostics so far. This failure is mainly due to the gap that separates biomarker discovery from clinical use; in particular, the rate limiting step in the generation of new diagnostics is represented by biomarker verification. To speed up this step, it is necessary to develop a biomarker verification technology that combines high sensitivity, high throughput, precision and specificity with the use of little amount of plasma.

The speaker presented a new approach for candidate biomarker validation, based on a mass spectrometry-based assay called SISCAPA. In this approach, anti-peptide antibodies immobilized on magnetic beads are used to enrich specific peptides together with spiked stable-isotope-labeled variants of the same sequence. After elution from the beads, on-line electrospray mass spectrometry is used to quantify the natural and labeled peptides using a Multiple Reaction Monitoring (MRM)-based quantitation method.

Clearly, this approach aims at the verification of pre-determined analytes, but it could be theoretically applied to biomarker discovery too, if the analysis is extended to selected proteotypic peptides representative of the complete human proteome (20,500 protein-coding genes).

My personal highlights concerning the general theme of the meeting “integration and proteome completion” give only a rough hint of the various and interesting issues addressed by the different speakers. The participants clearly enjoyed the conference topics and also the amazing Siena atmosphere. Moreover, the limited number of delegates allowed a relaxed and friendly atmosphere for scientific discussions.

Mara Colzani, Lausanne, CH