24.1

Proteomics Databases: Navigating the Archipelago

Henning Hermjakob

Swiss-Prot, European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SD, United Kingdom

Proteomics data is much more diverse than genetic data, important data types include for example, 2D gel images, mass spectra, and protein interactions. For each data type, the publicly available data is often distributed over several databases, journals' and authors' websites.

We will give an introduction to navigation in the proteomics data archipelago, zooming in from an overview map of data types, database contents and features to a more detailed exploration of protein-protein interaction databases. We'll point out treasure troves of high quality data and harbours of friendly user interfaces or what might await you after password-protected walls. In particular, we'll focus on bridges of interoperability between islands and possibilities of building new bridges of data integration.

25.1

Protein Chips and Bioinformatics: Essential Tools for Biomarker Discovery

Daniel W. Chan

Professor of Pathology, Oncology, Radiology and Urology; Director, Clinical Chemistry Division; Director, Biomarker Discovery Center; Johns Hopkins Medical Institutions, Baltimore, Maryland 21287-7065

The discovery of serum biomarkers is important in the diagnosis of diseases such as cancer. High throughput proteomic techniques, such as protein chips, coupled with the advances in bioinformatics would allow the identification of biomarkers useful for clinical application. One example is the surface-enhanced laser desorption-ionization (SELDI) time of flight mass spectrometry (TOF-MS). Like many other types of high-throughput expression data, protein array data are often characterized by a large number of variables (the mass peaks) relative to a small sample size (the number of specimens). In analyzing such data to screen for disease-associated biomarkers, it is important to extract as much information as possible from a limited number of samples and to avoid selecting biomarkers whose performances are influenced mostly by non-disease related artifacts in the data. A software package was developed by our group to compute and rank the contribution of each individual peak towards the optimal separation of two diagnostic groups. Several cases of cancer biomarkers discovery from serum or plasma will be shown.

24.2

A Tutorial on Systematic Proteomics Analyses

Soeren Schandorff

MDS Proteomics, Odense, Denmark

Technology developments in proteomics make possible systematic large-scale studies of protein interactions, differential expression, post-translational modifications and other pathway related information enabling challenging systems biology approaches. Bioinformatics is a fundamental component in the systems biology approaches for the organizing, analysing and display of these rather complex proteomics data. One additional element to this has been the revealing of the prevalence of alternative splicing which besides giving rise to higher complexity of the proteome also have added a serious complexity aspect to the distinction of information with respect to specific isoforms, which is currently rather poorly handled. This fundamental role of bioinformatic affects our way of thinking about proteomics and bioinformatic decisions partly govern on which terms we can compare different proteomics information and how. How can one be inclusive without drowning in information for which there is no common base for comparison? Starting from current large-scale proteomics analyses undertaken at MDS Proteomics and other proteomics companies an introduction will be given to the general concepts and approaches in tightly integrated bioinformatic solutions for the management and analyses of results from systematic proteomics studies.

25.2

Proteomic Analyses Using a Fully Automated Chip-based Nanoelectrospray System

Colleen K. Van Pelt[1], Sheng Zhang[1], Xian Huang[1], Gary Williams[2], and Mark H. Allen[2]

[1]Advion BioSciences, Ithaca, New York; and [2]Advion BioSciences, Ltd., Norwich, United Kingdom

A fully automated nanoelectrospray system used in combination with mass spectrometry has recently been developed. The NanoMate is a robotic infusion device that holds a 96-well plate, a rack of 96 disposable, conductive pipette tips, and an ESI Chip. The chip is a fully integrated monolithic electrospray device that consists of a 10 × 10 array of nozzles etched from the planar surface of a silicon wafer using standard semiconductor processes. This automated nanoelectrospray system offers low sample consumption and enhanced ionization resulting in higher sensitivity, similar to conventional pulled-capillary nanoelectrospray. However, the automated system offers a simple, one-time spray optimization for a set of 96 samples, no carryover between samples, and enhanced spray stability.

To demonstrate the capability of the NanoMate for proteomic analyses the results from several studies will be discussed including the identification of proteins from 2D gels of *E. coli* crude cell extract, non-covalent interactions between the endoglucanase catalytic domain of the *T. fusca* bacterium and cellopentose, protein quantification of standard proteins using isotope-coded affinity tags (ICAT), and phosphorylated peptide analysis using an affinity chromatography column charged with ferric ions. Results from off-line collected liquid chromatography fractions from *E. coli* crude cell extracts will also be shown.

25.3

A Sample Preparation Platform Technology for Proteomic Applications

M. Walid Qoronfleh

Perbio Sciences, Milwaukee, Wisconsin 53202-1009

Proteomics is having a profound impact on biological research. In this post-genomics era, there is a need to develop reliable methods to quickly and conveniently analyze multiple protein samples. Equally important, pre-fractionation of complex protein mixtures and enrichment of desired protein fractions are critical to proteome analysis. To this end, we have developed SwellGel™ Technology, a versatile and novel affinity chromatography platform. This technology utilizes a selected chromatography media that is dehydrated to form uniform aggregates. The discs will instantly rehydrate upon addition of the protein sample where the researcher need only perform a few simple steps to recover their protein/fraction of interest. Discs can be made in a variety of sizes (resin volume 25 μl–3 ml) dispensed in various formats (384-, 96-, 48-, and 24-well microplates or columns) and with different ligands attached to the matrix. SwellGel™ technology has been applied to affinity purification systems such as glutathione for GST-tagged proteins, metal chelates for His-tagged proteins, avidins for biotinylated molecule isolation, and proteins A/G for antibody purification. Data will be presented applying SwellGel discs to protein desalting/enrichment, the removal of abundant proteins from human serum- like albumin and immunoglobulin, and the isolation of phosphoproteins and phosphorylated peptides. The SwellGel™ format provides these chromatography systems with additional convenience, consistency, stability, and exceptional flexibility when compared to standard wet chromatography resins. These key features make SwellGel™ technology a significant advancement facilitating the sample handling needs for proteomic research.

25.4

Prioritising the Proteome: Identifying Pharmaceutically Relevant Targets by Linking Sequence to Function Through 3D Structure

John Overington

VP Drug Discovery, Inpharmatica, 60 Charlotte Street, London W1T 2NU, United Kingdom

Inpharmatica is focussed on the application of protein 3-D structural data to the identification of distantly related homologs of pharmaceutically important gene families. We have developed a combined bioinformatics and chemoinformatics platform that significantly improves the yield and quality of functional annotation and correlates this annotation with an estimation of druggability. We will describe some specific examples of the application of this technology to the target discovery process.

25.5

From Gene Products to Disease Phenotype: Antibody Chips for Predictive Medicine in Cancer

Moncef Jendoubi, Christine Chavany, Jean-Francois Martini, Yue Xu, and Rosaura Valle

Milagen Inc.

Characterizing the proteome has become essential to understanding phenotype and thus disease, leading in turn to potentially finding new and better ways to cure disease. Milagen has developed a unique antibody-based approach termed ANTIBIOMIX™ to directly correlate gene products to disease. A major feature of ANTIBIOMIX™ is the high-throughput generation of polyclonal antibodies in a relatively short period of time and at a reasonable cost. Milagen will complete its 100,000 polyclonal antibody collection against human gene products by the end of 2002.

To directly identify proteins that underlie disease, Milagen antibody collections are applied in multiplex format to large collections of clinical samples, using our proprietary matrix protein array technology. Our technology can be applied to tissues and biological fluids. Based on differential profiles between matched normal and disease samples, we have thus identified several cancer-related targets, in major cancer types cancer, as well as secreted molecules in sera. We intend to apply this technology to other major human diseases.

The identification of cancer specific targets in body fluids leads to the development of a new generation of diagnostic products, namely antibody chips for predictive medicine, applicable to early detection, disease progression, metastatic indications, toxicity profile, and therapy response. Secreted biomarkers are candidates for the development of therapeutic products as well.

ANTIBIOMIX™ is bridging the gap between genotype (gene sequence) and phenotype (diseases state), leading to the rapid identification of a variety of disease specific proteins, and to the validation of novel targets for diagnostic and therapeutic applications.

25.6

From Cell to Gel to Well: A Modular Automated Platform for Proteomics Assays

Christoph Eckerskorn

Tecan-Munich, Feldkirchner Strasse 12a, 85551 Kirchheim, Germany

Current traditional proteomics techniques can be insufficient for many researcher's needs in regards to: *sensitivity* (e.g., difficulties in detecting unknown, low abundant proteins), *reproducibility* (e.g., lack of robust, consistent methods for protein fractionation and 2D-PAGE), *throughput* (lack of automation systems and methods).

To help address these critical needs in Proteomics, Tecan has developed an automated modular platform. The complete platform consists of: a *Cell Maintenance System* (automation of cell culture procedures to insure cell preparation and treatment), a *Free Flow Electrophoresis System* (a novel semi-preparative, charged based, liquid separation technology for fractionation of cells, cell organelles and complex protein mixtures), a *2D-PAGE System* (a fully automated system from IEF, Gel casting, SDS-PAGE, and Staining), and a *Protein Processing System* (consisting of a spot picking device followed by an in gel digestion and an interface to mass spectrometry). The instrumentations can be integrated in various combinations (e.g., *Free Flow Electrophoresis* combined directly with *LC-MS*) and allows researchers to construct a customized platform for different proteomics applications. The current study presents protein fractionation and identification data generated with the automated platform with samples from human serum, yeast and liver.

In the current study, special emphasis will be placed on the use of the combination of *Free Flow Electrophoresis* with classical *2D-PAGE*. This technology is capable of overcoming one of the biggest challenges for proteome research: the reduction of the complexity of the protein samples by powerful early fractionation.