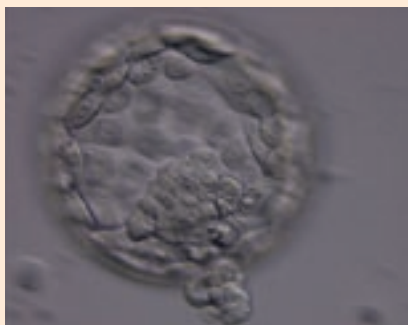


### Human embryo proteomics

Studies in which human embryos are destroyed are very controversial these days, so researchers have shied away from conducting proteomics analyses on embryos. But Mandy Katz-Jaffe and co-workers at the Colorado Center for Reproductive Medicine have taken an initial step toward characterizing the proteome of human pre-implantation blastocysts. The researchers say that such knowledge could shed light on myriad developmental processes in addition to interactions that occur between the embryo and the maternal uterine lining.

One roadblock facing proteomics scientists has been the lack of available human embryos for study. Because of the limited sample supply, Katz-Jaffe and co-workers developed a protocol to analyze the proteomes of individual embryos. Surplus cryopreserved embryos that had been donated by in vitro fertilization patients were thawed and cultured to the blastocyst stage. The embryos were classified as early, expanded, or degenerating. A total of 24 embryos



MANDY KATZ-JAFFE

**Uncharted territory.** Researchers performed proteomics studies on human embryos that were at the blastocyst stage.

were lysed and individually stored at  $-80^{\circ}\text{C}$ . Thawed lysates were then analyzed by SELDI TOFMS with strong-cation-exchange protein chips.

Although some variability existed among embryos within the same morphological classification, the researchers also detected a few significant differences between embryos in the different groups. For example, two potential biomarkers were up-regulated in expanded blastocysts as compared with early blastocysts. Katz-Jaffe and colleagues found a total of six potential biomarkers that distinguish degenerating embryos from developing embryos at the early or expanded blastocyst

stages. They identified candidate biomarkers, and several of these peptides and proteins—such as parathyroid-hormone-related peptide and heparin-binding epidermal-growth-factor-like growth factor precursor—are known to be involved in embryogenesis. According to the researchers, the identifications are preliminary, and additional studies are necessary to validate these results. (*Fertil. Steril.* **2006**, *85*, 101–107)

### An interactome map for humans

As a step toward a systemwide understanding of how organisms function, researchers have constructed maps of protein–protein interactions (PPIs) for the yeast, worm, and fly. Now, Akhilesh Pandey and co-workers at the Institute of Bioinformatics (India), Johns Hopkins University, and the University of Würzburg (Germany) have taken the next leap: developing an interactome map for humans. Surprisingly, only a few human interactions were common to the interactomes of the model organisms.

The human interactome map was based on information culled from several databases, such as the Human Protein Reference Database (known as HPRD), the Biomolecular Interaction Network Database (known as BIND), and the Database of Interacting Proteins (known as DIP). In total, 25,464 human PPIs were mapped.

Conventional wisdom holds that PPIs across related organisms are conserved. To test this idea, Pandey and co-workers compared the human interactome map with maps containing 16,069 yeast, 5625 worm, and 24,587 fly PPIs. Only 16 PPIs

were common to all 4 organisms, and 42 PPIs were common to human, worm, and fly. The worm and fly data were derived from two-hybrid experiments, which can yield many false positives. So in the next analysis, the researchers considered only bait and prey proteins from the worm and fly that were similar to orthologs of human proteins. In this case, 42 of 288 possible common PPIs were shared among human, worm, and fly.

Some scientists have suggested that essential genes are the most highly interconnected proteins in an organism. Pandey and co-workers say that this relationship was observed and that their data were statistically significant for yeast and humans. However, they also say that the difference is not dramatic and that one cannot predict whether a gene is essential on the basis of connectivity.

Finally, Pandey and co-workers used the Online Mendelian Inheritance in Man database to investigate the relationship between gene interaction networks and disease. In the analysis, genes that were involved in the same types of diseases or in diseases that occur in the same organ were highly likely to inter-

act with each other. According to the researchers, the protein and gene interactions identified in this study could be a springboard for systems biology investigations. (*Nat. Genet.* **2006**, *38*, 285–293)

### Metabonomics for epidemiology

The goal of the International Study of Macro/micronutrients and Blood Pressure (INTERMAP) project, launched in 1996, is to determine the relationship between diet and blood pressure. To get an idea of the experimental and biological variation that will be encountered in the study, Marc-Emmanuel Dumas, Elaine Holmes, and co-workers at Imperial College London, Shiga University of Medical Science (Japan), the Chinese Academy of Medical Sciences, and Northwestern University used  $^1\text{H-NMR}$  spectroscopy to perform an initial metabonomics analysis on INTERMAP samples.

Urine samples collected from subjects in Aito Town (Japan), Chicago, and Guangxi (China) were studied. At the collection sites, some of the samples were split into two aliquots, which were assigned different identification numbers so that a double-blind quality

## TOOLbox

### Making a SPLASH

Massive amounts of data are generated in high-throughput proteomics experiments. To manage, search, and store this information, Maxey Chung and co-workers at the National University of Singapore have developed the systematic proteomics laboratory analysis and storage hub (SPLASH).

SPLASH consists of three modules. With the maintenance module, researchers can enter their results and descriptions of their analytical methods and samples. Researchers can search entries in SPLASH by several parameters, including the accession number and protein name with the search module. In addition, users can click directly on a gel image to search for more information. The mining module allows researchers to analyze the data with Gene Ontology terms or Kyoto Encyclopedia of Genes and Genomes (known as KEGG) biochemical pathway information. SPLASH is available at <http://oncoproteomics.nus.edu.sg/splash>. (*Proteomics* 2006, 6, 1758–1769)

### PRISM

Richard Smith and colleagues at the Pacific Northwest National Laboratory (PNNL) have developed a computational platform called the proteomics information storage and management (PRISM) system to handle accurate mass and time (AMT) tag data. PRISM is composed of two major parts. The data management system (DMS) collects and tracks sample metadata and raw spectra. Each storage server within the DMS holds 2–6 terabytes of information. The DMS also includes manager programs that automatically process AMT data. The mass and time tag system combines results from different AMT experiments. An alignment algorithm corrects the retention times of identified species and normalizes the values. Peptides that are observed in several experiments are given an idealized mass and time tag; this tag allows scientists to increase their throughput and to have higher confidence in the data. Although PRISM was originally developed for PNNL use, the researchers plan to make the platform more widely available to researchers. (*Proteomics* 2006, 6, 1783–1790)

control analysis could be conducted. A hierarchical clustering approach was applied to the  $^1\text{H-NMR}$  spectral data to assess the degree of similarity among the samples. Split samples derived from the same original specimen should cluster together. In total, 98% of the Japanese, 96% of the U.S., and 71% of the Chinese split samples were correctly grouped. To investigate the reason for the low detection rate in the Chinese samples, a second, unrelated set of samples was collected from people in Beijing. After those samples were split, 93% were aggregated correctly. Thus, the researchers say that the Guangxi specimen variability was probably due to inconsistencies in sample handling. Overall, physicochemical variation (e.g., pH), biochemical degra-

dation, and analytical instrumentation variation caused errors in split-sample identifications.

The coefficients of variation for the metabolic signals were 0–10%. The researchers say that these results demonstrate that  $^1\text{H-NMR}$  spectroscopy is more reproducible than other instrumentation, such as a TOF mass spectrometer designed for metabolic studies and gene microarrays. In addition, the researchers were able to discriminate spectra from Japanese, American, and Chinese samples with statistical analyses. Major metabolites were also identified in the study. These data suggest that metabolomics methods are well suited to large epidemiological studies. (*Anal. Chem.* 2006, 78, 2199–2208)

### Which proteins are involved in differentiation?

The cells in our bodies are not all the same. Some cells have long protrusions; others are smaller and more compact. Cell functions can vary, too. For example, neurons help us to sense our surroundings and to think, whereas muscle cells help us to move through our environment. What makes a neuron look and function like a neuron, as opposed to a muscle cell?

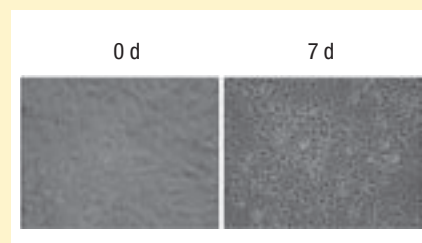
Arndt Rolfs and colleagues at the University of Rostock, Ernst Moritz Arndt University, DECODON GmbH (all in Germany), ReNeuron, Ltd. (U.K.), and Rzeszow University (Poland) have made an initial attempt to answer this question by using proteomics methods.

A stable human fetal neural stem cell (NSC) line, which can be coaxed to become a group of neurons, astrocytes, or oligodendrocytes *in vitro*, was analyzed in the study. The researchers ran 2DE gels of extracts from undifferentiated (proliferating) cells and from cells that had undergone differentiation for 4 or 7 d. All three gel images were

overlaid to produce a reference master gel of NSC proteins. Of the 956 spots that were mapped, 402 spots contained 318 unique proteins that were identified by MALDI TOFMS. The proteins were classified on the basis of functional information obtained from the Gene Ontology database and the ExPasy server.

NSCs that are differentiating undergo changes in shape that correlate with changes in protein expression.

When Rolfs and colleagues compared the protein patterns of proliferating cells with the patterns obtained from differentiating cells, they found 146 protein spots that varied >1.5-fold in intensity. Of these spots, about half were more intense in differentiating cells and about half were less intense. The researchers then validated the expression levels of four proteins identified in the study. Quantitative western blots confirmed that differentiating cells express more transgelin-2 and peroxiredoxin 1, but less peroxiredoxin 4 and proliferating cell nuclear antigen protein than proliferating cells. (*Proteomics* 2006, 6, 1833–1847)



**The change.** The appearance of proliferating NSCs (0 d) is different from that of NSCs that are becoming specialized (7 d). (Adapted with permission. Copyright 2006 Wiley-VCH Verlag GmbH & Co.)

## PBDEs and the developing brain

Neonatal exposure to the brominated flame retardant 2,2',4,4',5-pentabromodiphenyl ether (PBDE-99) during brain development could have long-lasting neurotoxic effects. Henrik Alm and colleagues at Uppsala University (Sweden) used a proteomics-based approach to study the effects of exposure to PBDE-99 on the striatum and hippocampus of the neonatal mouse brain during a critical stage of brain growth. Using difference in-gel electrophoresis, they found that 40 protein spots in striatum samples and 56 protein spots in hippocampus samples had significantly altered levels.

Proteins involved in neurodegenera-

tion and neuroplasticity were differentially regulated in the striatum, and proteins involved in metabolism and energy production were differentially regulated in the hippocampus. The researchers say that the combination of effects during this critical stage may contribute to late behavior outcomes.

The identified proteins reflect the results of early exposure to PBDE-99 and could be used as biomarkers of developmental neurotoxicity. The researchers suggest that a similar proteomics-based approach could be used to investigate the neurotoxic effects of exposure to other organohalogen compounds during brain development. (*Environ. Health Perspect.* 2006, 114, 254–259)

## Ultramicroarrays for limited sample quantities

Typically, clinical and forensic samples are available only in small volumes. But many known biomarkers are present in low concentrations in these limited samples. To detect these rare proteins, S. Nettikadan and colleagues at BioForce Nanosciences Inc., Iowa State University, and Des Moines University have developed ultramicroarrays with detection limits in the attomole range. The researchers could detect proteins from as few as four cells in some experiments.

Whereas spots on conventional microarrays are 100–300  $\mu\text{m}$  in diameter, spots on ultramicroarrays are much smaller, usually 1–20  $\mu\text{m}$  in diameter. Smaller spots increase microarray sensitivity, according to the researchers. To produce ultramicroarrays, antibodies are immobilized on gold-coated silicon surfaces by an amine-reactive chemistry. Proteins bound to antibodies are detected by standard antibody sandwich methods and fluorescence microscopy.

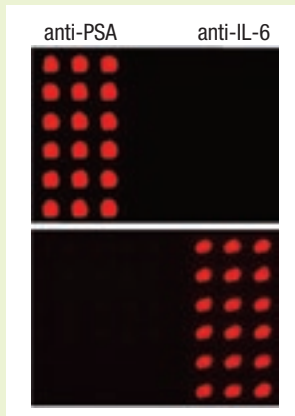
To test specificity, a set of spots

containing antibodies to prostate-specific antigen (PSA), a marker for prostate cancer, was printed alongside a set of spots containing interleukin-6 (IL-6), a possible marker for prostate cancer. These PSA/IL-6 ultramicroarrays were

incubated with purified PSA or IL-6. Only the anti-PSA spots were labeled when purified PSA was used, and only anti-IL-6 spots were labeled when purified IL-6 was used.

In other experiments, Nettikadan and colleagues tested ultramicroarrays with anti-PSA spots only. With the new array format, the researchers could detect PSA secreted by as few as 4 prostate cancer cells within a 24-h period. They could also detect PSA secreted by 100 cells in <3 h. Intracellular PSA from as few as six cells

was easily detected. The researchers say that ultramicroarrays are more sensitive than other reported protein microarray formats and could be used to screen samples such as those obtained by laser-capture microdissection. (*Mol. Cell. Proteomics* 2006, doi 10.1074/mcp.M500350-MCP200)



**Bright lights.** Ultramicroarrays made with anti-PSA are specific for PSA, and ultramicroarrays made with anti-IL-6 are specific for IL-6. (Adapted with permission. Copyright 2006 American Society for Biochemistry and Molecular Biology.)

## Integrating heterogeneous data sets

Many biological databases exist, but few tools are available for relating individual entities in one database to entities in another. To retrieve interrelated information, such as how a protein functions in the presence of other proteins, Golan Yona and Aaron Birkland of Cornell University have developed Biozon, a system for unifying heterogeneous biological data into a nonredundant dataset. Biozon takes into account the overlapping nature of data so that the broader context of biological entities can be gleaned. The system not only warehouses data; it also computes and stores derived data such as similarity relations and functional predictions. It also detects conflicts between source databases. Currently, Biozon holds >100 million biological documents and 6.5 billion relations between them. The system can be easily updated as new data become available. The database can be accessed at <http://biozon.org>. (*BMC Bioinformatics* 2006, 7, 70)

## Combined database of network interactions

In systems biology, researchers must combine numerous databases into one management scheme to obtain biological insight. So Michael Baitaluk and colleagues at the University of California, San Diego; Keck Graduate Institute; and Claremont Graduate University developed PathSys, a graph-based system for integrating such large amounts of information into a database of interaction networks. The system integrates different networks, including protein–protein interactions, metabolic pathways, and genetic interactions, into one system that can be used to address questions relevant to systems biology. A graph query language and engines for evaluating queries make it possible for the system to process complex input from users. Thus far, Baitaluk and colleagues have used PathSys to integrate >14 curated data sources for the yeast *Saccharomyces cerevisiae*; however, they say that the system could be applied to any biological model. The PathSys project homepage is at <http://brak.sdsc.edu/pub/BiologicalNetworks/PathSys>. (*BMC Bioinformatics* 2006, 7, 55)

## TOOLbox

### ProteomeCommons.org JAF

P. C. Andrews and co-workers at the University of Michigan have developed the Java analysis framework (JAF), a library of freely available open-source code. The researchers created JAF as an application programming interface (API) to speed the development of bioinformatics tools. The JAF API components include configurable libraries of atomic masses, amino acids, and known modifications to amino acids. In addition, tools to determine isotopic distribution and the presence of single nucleotide polymorphisms are available. The user tools include HTML references of atomic masses, amino acids and their modifications, and amino acid combinations and mass shifts. An online peptide  $m/z$  and  $pI$  calculator predicts masses,  $pI$  values, and proteotypic fragments for peptides. The online residue combination calculator predicts all of the possible combinations of amino acids for a particular  $m/z$  value. Researchers can access JAF at [www.proteomecommons.org](http://www.proteomecommons.org). (*Bioinformatics* **2006**, *22*, 632–633)

### MasSPIKE

A bottleneck in proteomics experiments is spectrum interpretation. This process is a particular challenge when analyzing complicated spectra generated by FTMS instruments in top-down experiments. So Peter O'Connor and Parminder Kaur developed a suite of algorithms called mass spectrum interpretation and kernel extraction (MasSPIKE) that convert high-resolution spectral data into monoisotopic peak lists. MasSPIKE models noise across the spectrum, identifies isotopic clusters, determines the charge states, and separates overlapping isotopic distributions. In addition, the programs choose the isotopic peaks and align the experimentally observed and theoretically derived distributions. If a known protein is fragmented, then  $b$  and  $y$  ions are calculated and assigned automatically. O'Connor and Kaur tested the algorithms on a spectrum of bovine carbonic anhydrase. They plan to include MasSPIKE in the Boston University Data Analysis (known as BUDA), an open-source software package available at [www.bumc.bu.edu/ftms](http://www.bumc.bu.edu/ftms). (*J. Am. Soc. Mass Spectrom.* **2006**, *17*, 459–468)

### Top-down proteomics on an Orbitrap mass spectrometer

In the top-down proteomics approach, researchers obtain better sequence coverage than with bottom-up approaches because whole proteins are analyzed. The problem is that only expensive, high-maintenance FTICR instruments have enough resolving power and mass accuracy to adequately analyze the protein ions. But Matthias Mann and colleagues at the Max Planck Institute of Biochemistry (Germany) and the University of Southern Denmark demonstrate that the linear ion trap (IT)–Orbitrap hybrid mass spectrometer, which is less expensive and requires less maintenance, performs as well as or better than a linear IT–FTICR instrument for top-down experiments.

Mann and colleagues ran the new IT–Orbitrap mass spectrometer through a battery of tests. When they tested for sensitivity, they detected as little as 50 fmol of cytochrome *c* with the instrument. Although other scientists have reported higher sensitivities with FTICR devices, they have done so only with special methods. In addition, a resolution of 100,000 (at a 1.8-s measure-

ment time) is possible with the IT–Orbitrap. This resolution is lower than those achieved on FTICR instruments, but the resolution remained higher across a given mass range with the IT–Orbitrap. Furthermore, the absolute mass accuracies for each of six protein standards ranged from 0.92 to 2.8 ppm with this new instrument. The researchers say that mass accuracy <10 ppm is rarely reported for FTICR machines. Finally, the IT–Orbitrap was used to detect the phosphorylated forms of three proteins.

Little fragmentation was observed with nozzle–skimmer dissociation in the IT–Orbitrap instrument, so  $MS^2$  was routinely performed by collision-induced dissociation in the linear IT. All six proteins tested were unambiguously identified with  $MS^2$  fragments generated in the IT–Orbitrap. Fewer  $MS^2$  fragments were identified with data from an IT–FTICR instrument. Mann and colleagues also obtained interpretable spectra from  $MS^3$  and  $MS^4$  scans in the IT–Orbitrap. In addition, the researchers fragmented peptide-like  $MS^2$  fragments in a simulation of a bottom-up approach in the IT–Orbitrap. (*Mol. Cell. Proteomics* **2006**, doi 10.1074/mcp.T500042-MCP200)

### ICPMS for proteomics

Protein phosphorylation is an important posttranslational modification that is involved in the regulation of many processes. To quantitatively assess phosphorylation, Wolf Lehmann and colleagues at the German Cancer Research Center and the University of Cologne (Germany) developed two methods based on inductively coupled plasma MS (ICPMS).

In the  $\mu$ LC–ICPMS approach, proteins are separated by 1DE gels and in-gel digested. The peptides are then fractionated by reversed-phase capillary LC and analyzed by ICPMS. In the laser ablation (LA)–ICPMS approach, proteins are run on 1DE gels and blotted onto membranes. These intact proteins are then analyzed by LA–ICPMS, in which parallel line scans are made across the membranes and color-coded 2D images are produced (imaging LA–ICPMS). Phosphorus ( $^{31}\text{P}$ ) is monitored to determine the degree of phosphorylation, whereas sulfur ( $^{34}\text{S}$ ) is monitored to determine the overall protein content in both methods.

The researchers evaluated the feasibility of the  $\mu$ LC–ICPMS method by analyzing ovalbumin. The experimentally obtained P/S ratio closely matched the value reported in the literature. To evaluate LA–ICPMS for phosphorylation studies, a mixture of ovalbumin and three other proteins was analyzed. The presence of 2 bands on the  $^{31}\text{P}$  image correctly indicated that 2 proteins were phosphorylated, but the presence of 3 bands on the  $^{34}\text{S}$  image incorrectly indicated that only 3 proteins were in the sample. The fourth protein, myoglobin, was too faint to be seen on the  $^{34}\text{S}$  image, however, because it only contains two S atoms per molecule.

Both ICPMS methods were applied to the proteomes of the bacterium *Corynebacterium glutamicum* and the mouse *Mus musculus*. Although many  $^{31}\text{P}$  signals were detected in the mouse sample, few were detected in the bacterial sample. The minimal phosphorylation seen in the bacterial sample is consistent with the fact that prokaryotes contain fewer kinases than eukaryotes. (*Anal. Chem.* **2006**, *78*, 1987–1994)