

SCIENTIFIC SESSION SPEAKER ABSTRACTS

S1

The Human Genome and Beyond

R. H. Waterston; University of Washington, Seattle, WA, United States.

The sequence of the human genome is now essentially complete, with more than 99% of the euchromatic sequence available in the public domain for unrestricted use. The project catalyzed enormous gains in sequencing efficiency and capacity. As a result, genome sequence is now available for several experimentally important eukaryotes and a host of prokaryotes, with many more genomes underway and still more planned in the coming years.

The task now shifts to understanding the content of the human and other genomes. An immediate challenge is the definition of the 'parts list' – the functional elements of the genome. The analyses of the genomes from several yeasts, two nematodes and mouse and human illustrate the power of comparative sequencing in contributing to this task. These comparisons exploit the fact that purifying selection maintains functional sequence, whereas nonfunctional sequence accumulates change from neutral drift. With pairwise comparison, significantly improved gene sets can be obtained, while with comparisons of more genomes at an appropriate evolutionary distance, regulatory sequences begin to emerge. With the capacity to generate several mammalian genome sequences in the next few years, an increasingly complete human 'parts list' will be in hand.

S2

Mechanistic studies and anti-HIV therapeutic applications of RNAi

J. Rossi; Beckman Research Institute of the City of Hope, Duarte, CA, United States.

RNAi is a powerful, sequence specific approach for interfering with gene expression. In order to better understand the biochemistry of this mechanism, we have identified a specific siRNA binding protein using a Northwestern hybridization approach. Using this approach we identified the well known RNA binding protein La. Subsequent experiments demonstrated that La is the RNA binding component of a multi-protein complex that unwinds siRNAs in an ATP dependent fashion. These complex hands off the antisense strand to the RNAi induced silencing complex, or RISC, which is the functional entity that facilitates targeted RNA destruction. An interesting corollary to this finding is the independent discovery that a positive strand plant virus encoding protein, p19 competes with La for siRNA binding both in cell extracts and in vivo. P19 can block RNAi in mammalian cells, suggesting that human positive strand RNA viruses could use similar mechanisms to overcome RNAi. Thus, by understanding the mechanisms by which viruses can inhibit RNAi, it may be possible to develop therapeutic strategies to overcome this inhibition. We are currently evaluating the biological consequences of inhibiting RNAi in mammalian cells using micro arrays.

We have also been exploring the use of expressed siRNAs for the treatment of HIV-1 infection as well as applications in cancer. Our approach has involved the use of siRNAs expressed from Pol III and Pol II promoters, some of which are incorporated in lentiviral vectors. We have transduced such vectors encoding anti-HIV-1 siRNA genes into hematopoietic progenitor cells and shown that the expressed siRNAs can

potently inhibit HIV-1 infection in macrophages and T-lymphocytes derived from the precursors.

In addition to using siRNAs to knockdown RNA levels, we have begun to investigate the potential for siRNA directed methylation of target genes as a method for gene silencing at the level of DNA. Short hairpin RNAs targeting CpG islands in the RASSF1A gene were expressed from the human U6 promoter in stably transfected HeLa cells. The target RASSF1 gene is completely unmethylated in HeLa cells but is heavily methylated and silenced in several tumors. Using this model system we demonstrate that siRNAs can direct sequence specific methylation of DNA in human cells. A reduced level of transcription accompanied methylation as well. With the exciting prospect of being able to direct DNA methylation with shRNAs (or siRNAs), it may be feasible to effect silencing of viral and cellular genes at the transcriptional level as well as the post-transcriptional level.

S3

Rapid Identification of Functional Genes by Libraries of Ribozymes and siRNAs

K. Taira; The University of Tokyo, Tokyo, Japan.

RNA interference (RNAi) is a phenomenon whereby expression of an individual gene is specifically silenced by the introduction of a double-stranded RNA (dsRNA) whose sequence is homologous to that of the gene in question. The generation of a small interfering RNA (siRNA) expression library directed against the entire human genome is a project that requires solutions to many difficult technical problems. We have solved many of these problems, including the development of genetically stable and highly active siRNA expression vectors, a procedure for selection of favorable target sites, and an efficient and inexpensive procedure for constructing a siRNA expression library. Libraries were also made of hammerhead ribozymes with randomized binding arms, in addition to U6- and tRNA-driven siRNA libraries. Then, the libraries were introduced into cells either by transfection or by viral vectors. This procedure made it possible to readily identify the relevant genes associated with phenotype in the apoptosis, cancer metastasis, and/or cell differentiation pathways. This application of the library represents a simple and yet powerful method for identification of functional genes associated with specific phenotypes in the post-genome era. Importantly, the identified functional genes originated not only from the coding region but also from the noncoding region.

REFERENCES

1. Zhou, D.-M. and Taira, K. *Chem. Rev.*, **98**, 991-1026 (1998);
2. Kuwabara, T. *et al. Nature Biotechnology*, **16**, 961-965 (1998);
3. Kawasaki, H., *et al. Nature*, **393**, 284-289 (1998);
4. Kuwabara, T. *et al. Mol. Cell*, **2**, 617-627 (1998);
5. Kuwabara, T. *et al. Proc. Natl. Acad. Sci. USA*, **96**, 1886-1891 (1999);
6. Tanabe, T. *et al. Nature*, **406**, 473-474 (2000);
7. Warashina, M. *et al. Proc. Natl. Acad. Sci. USA*, **98**, 5572-5577 (2001);
8. Kato, Y. *et al. J. Biol. Chem.*, **276**, 15378-15385 (2001);
9. Takagi, Y. and Taira, K. *J. Am. Chem. Soc.*, **124**, 3850-3852 (2002);
10. Tanaka, Y. *et al. J. Am. Chem. Soc.*, **124**, 4595-4601 (2002);
11. Suzumura, K. *et al. J. Am. Chem. Soc.*, **124**, 8230-8236 (2002);
12. Kawasaki, H. *et al. Nature Biotechnology*, **20**, 376-380 (2002);
13. Kawasaki, H. and Taira, K. *EMBO Rep.*, **3**, 443-450 (2002);
14. Onuki, R. *et al. Proc. Natl. Acad. Sci. USA*, **99**, 14716-14721 (2002);
15. Miyagishi, M. and Taira, K. *Nature Biotechnology*, **20**, 497-5000 (2002);
16. Wadhwa, R. *et al. EMBO Rep.*, **4**, 595-601 (2003);
17. Suyama, E. *et al. Cancer Res.*, **63**, 119-124 (2003);
18. Suyama, E. *et al. Proc. Natl. Acad. Sci. USA*, **100**, 5616-

5621 (2003); 19. Nelson, L. D. *et al. TARGETS*, **2**, 191-200 (2003); 20. Kawasaki, H. and Taira, K. *Nature*, **423**, 838-842 (2003); 21. Tanaka, Y. *et al. J. Am. Chem. Soc.*, in press; 22. Miyagishi, M. and Taira, K. *Oligonucleotides*, in press; 23. Onuki, R. *et al. EMBO J.*, in press.

S4

The Development and Application of RNAi as a Functional Genomics Tool

N. Caplen; NIH, Bethesda, MD, United States.

RNA interference (RNAi) is a sequence-specific post-transcriptional gene silencing (PTGS) mechanism triggered in somatic mammalian cells by RNA duplexes of 21 – 23 nucleotides (nts) called small interfering RNAs or siRNAs. siRNA can be introduced into mammalian cells in a number of forms including duplexes formed from chemically synthesized RNA oligonucleotides and from single stranded RNA transcripts that form a hairpin stem loop structure, where the stem region corresponds to the siRNA sequence. Short hairpin RNAs, as these are known have the advantage that these can also be expressed intracellularly from pol II or pol III promoters and can thus be cloned into plasmid and viral vectors. RNAi has been rapidly adopted as a functional genomics tool in a wide range of species and has been adapted to allow for the transient or stable knockdown of gene expression generation in cell lines and animals and has been developed for high throughput analysis of gene function in *Caenorhabditis elegans* and *Drosophila*. Several programs are in progress to generate siRNA sequences (both as synthetic siRNAs and as shRNAs) corresponding to most of the identified transcripts within the human genome. In order to facilitate large scale, high throughput functional genomics studies using RNAi, we have developed a microarray-based technology for highly parallel analysis. Specifically, siRNAs in a transfection matrix are arrayed on glass slides, overlaid with a monolayer of adherent cells, incubated to allow reverse transfection, and assessed for the effects of gene silencing by digital image analysis at a single cell level. Validation experiments with HeLa cells stably expressing GFP showed spatially confined, sequence-specific, time- and dose-dependent inhibition of green fluorescence for those cells growing directly on micro-spots containing siRNA targeting the GFP sequence. Microarray-based siRNA transfection analyzed with a custom-made quantitative image analysis system produced results that were identical to those from traditional well-based transfection, quantified by flow cytometry. Finally, to integrate experimental details, image analysis, data display and data archiving, we have developed a prototype information management system for high-throughput cell-based analyses. This RNAi microarray platform, together with ongoing efforts to develop large-scale human siRNA libraries, should facilitate genomic-scale cell-based analyses of gene function.

S5

Molecular Staging in Colorectal Cancer

S. A. Bustin; University of London, London, United Kingdom.

Conventional histopathological staging of primary colorectal cancers does not allow accurate prognostic stratification within a given tumour stage. Therefore, PCR-based assays are increasingly used to try and predict more accurately the likelihood of disease progression for the individual patient. Real-time reverse transcription PCR (RT-PCR) assays were used to detect and quantitate cytokeratin 20 (ck20), carcinoembryonic antigen (CEA) and guanylyl cyclase C

(GCC) mRNA in 149 lymph nodes (LN) from 17 patients with benign disease and 302 LN from 42 patients with colorectal cancer who had curative (R0) resections. None of the markers were specific, with ck20, CEA and GCC mRNA detected in 47%, 89% and 13% of 149 LN, respectively, from patients with benign disease. The sensitivity of all three markers was very high, with mRNA detected in 93%, 100% and 97% of 30 histologically involved LN, respectively. There was significant overlap in the mRNA levels of all three markers between histologically involved and uninvolved LN. There was no association between mRNA levels and distant recurrence (median follow-up: 3.94 years, range 3.35-5.12). We conclude that the use of molecular techniques to detect occult disease in LN may suffer from the same limitations as conventional methods. Instead, accurate prognostic stratification requires careful assessment of the likely metastatic potential of the primary cancer.

S6

Genomic Analysis of Ovarian Cancer

D. Ginzinger; UCSF Comprehensive Cancer Center, San Francisco, CA, United States.

Ovarian cancer is fifth most common killer of women in the United States, yet the death rate has not changed significantly over the last fifty years. This is due to the late stage in which the patient is typically diagnosed. As with all tumors, ovarian cancer is characterized by cellular growth beyond normal control and is associated with considerable genomic instability. A method pioneered at UCSF, called comparative genomic hybridization (CGH), has been instrumental for delineating the extent and scale of genomic aberrations recurrent in ovarian cancer. Following this paradigm, these recurrent genomic aberrations of DNA copy number changes are the underlying force driving the evolution of the tumor. Subsequently, these DNA copy number aberrations in turn drive the expression of genes responsible for the uncontrolled growth of the epithelial cells that constitute the majority of ovarian tumors. As part of a large study to perform detailed genomic analysis on 20 cell lines and up to 30 tumors, we used array CGH, Affymetrix arrays and real-time PCR analysis to measure the DNA copy number and gene expression levels of genes in regions of aberrant DNA copy number. We found very high concordance between DNA copy number changes and expression levels as well as very high concordance between the two methods of measuring expression levels. Details of the real-time PCR component will be presented on 50 genes measured in up to 50 samples as well as the resulting concordance with Affymetrix arrays.

S7

NMR-Based Structural Genomics

M. A. Kennedy, J. R. Cort, T. A. Ramelot, S. Ni; Pacific Northwest National Laboratory, Richland, WA, United States.

The Environmental Molecular Sciences Laboratory (EMSL) at the Department of Energy's Pacific Northwest National Laboratory (PNNL), Richland, Washington, USA, houses the High Field Magnetic Resonance Facility (HFMR) which contains 15 NMR spectrometers ranging from 300 MHz to 900 MHz in proton resonance frequency. The HFMR is an important resource for the Northeast Structural Genomics consortium (NESG) headed by Dr. Gaetano Montelione at Rutgers University, Piscataway, New Jersey, USA. The NESG is one of the National Institutes of Health's (NIH) Protein Structure Initiative (PSI) P50 Pilot Centers for structural

genomics. The PSI pilot centers are challenged to test the feasibility of various designs and strategies that might be adopted in the production stage of NIH supported PSI research. The participation of PNNL in the NESG provides for an interesting assessment of the feasibility of distributed data collection and structure determination as one paradigm for how production-scale structural genomics facilities might operate. We will discuss advantages and limitations in the use of national shared NMR resources as a component of production scale facilities. Functional annotation of protein structures is a natural part of the scientific process of any structural genomics effort. Based on our analyses, we will discuss the feasibility of NMR-structure-based functional annotation using the NESGC target structures determined at PNNL. Finally, results will be discussed for a study involving 100 protein targets from *Vibrio cholerae* designed to investigate the relative suitability of hypothetical protein targets for NMR and/or crystallography.

S8

Microarray Analysis of Alternative Splicing

H. Li, R. Dinakar, Y. Kwon, X. Fu; University of California, San Diego, CA, United States.

We have previously developed a microarray platform known as RASL (RNA Annealing Selection and Ligation) to profile alternative splicing (Nat. Biotech. 20:353-358, 2002). This technology is based on the use of an index oligonucleotide sequence, which is coupled with a pair of oligonucleotide sequences targeting a specific mRNA isoform. After the annealing reaction and solid phase selection, paired oligonucleotide sequences are ligated followed by PCR amplification using universal primers. The products are then hybridized to a universal microarray containing spotted index sequences. We previously tested the RASL technology in a small scale, and have now gradually increased the multiplexicity of the assay to profile more than 1000 mRNA isoforms. We are now applying the RASL technology to investigate alternative splicing events that are regulated in specific cell types or under control by specific splicing regulators. Specific examples of tissue-specific and regulated alternative splicing will be presented and their biological implications will be discussed.

Coupling with our effort in technology development, we are also constructing an alternative splicing database based on computer-assisted manual annotation. A series of bioinformatics tools for both mRNA isoform annotation and microarray data analysis are being developed. We will discuss the potential link and synergy between our efforts in database construction/array design and other array systems currently under development in both academia and industry.

S9

Discovery and tissue-specific monitoring of alternative pre-mRNA splicing events using ink-jet microarrays

J. M. Johnson; Rosetta Inpharmatics, Merck & Co., Kirkland, WA, United States.

To understand the biological roles, mechanism, and regulation of alternative splicing, systematic methods to monitor the splicing of exons into mature transcripts across different tissues and stages of development are needed. Microarrays offer a sensitive and parallel means to monitor transcript structure at high resolution and detect tissue-specific alternative splicing events. We have developed and optimized microarray design, sample amplification, and data analysis methods that allow tissue-specific detection of exon-intron edges, exon-exon

junctions, and alternative splicing events using ink-jet microarrays. These tools form a robust and scalable platform we have used for discovery and analysis of alternative splicing at exon-exon junctions in a systematic survey of human genes in over 50 tissues. We have used the results of these experiments to discover novel alternatively spliced isoforms of drug target genes and map the tissue distributions of known isoforms.

S10

Analysis of gene expression diversity at the level of alternative mRNA splicing using GeneChip® arrays

J. Hu, G. Lu, H. Wang, M. Cline, T. Clark, G. Mei, R. Wheeler, A. Williams, J. Blume; Affymetrix, Santa Clara, CA, United States.

Alternative mRNA splicing can expand the diversity of gene expression by generating multiple forms of mRNA and proteins with potentially different functions from a single gene in eukaryotes. Approximately 40-60% of human genes have been estimated to undergo alternative splicing. Some of these splicing events have been implicated in different diseases and aberrant biological functions. Therefore, it is essential to distinguish which splice variants are expressed in complete expression profiling. Exon-based splice variant expression analysis, in contrast to whole gene-based overall expression analysis, becomes the critical next step. We have developed an approach for alternative splicing analysis using GeneChip® oligonucleotide arrays. Previously, we have tested this approach using a human splicing test array containing 21 genes with known alternative splicing. Subsequently, we investigated the applicability of this approach to a larger number of genes with modified probe and chip designs. A new splicing test array was designed to detect alternative splicing of over 7000 mouse genes and was hybridized to cDNA targets from a number of mouse tissues and splice variant clones. We developed a high-throughput data mining and analysis strategy to identify alternative splicing based on training data from splicing variants of cDNA clones. Using this strategy, we could identify and sort out numerous alternative splicing events from various mouse tissues. We confirmed alternative splicing events of a few genes by TaqMan assays. Our studies indicate that alternative splicing can be detected at a large scale using this approach with GeneChip® oligonucleotide arrays. This approach will allow us to define more accurately the expressed genome and its functions.

S11

Profiling Alternative Splicing Using Microarrays

M. Ares, K. Srinivasan, C. Sugnet, L. Shiue, R. Centers, L. Edmondson, S. Fitzwater, J. Bryant, M. Smith, R. Loewen, Y. Mandel-Gutfreund; University of California, Santa Cruz, Santa Cruz, CA, United States.

Gene expression profiling methods have largely ignored the problem of alternative splicing. Yet the average mammalian gene produces 3 different mRNAs, in many cases coding for proteins of distinct function. In order to profile gene expression with sufficient resolution to capture information about alternative splicing, we have been building and testing oligonucleotide microarrays. Technological challenges and special data handling needs that have arisen as a consequence of these efforts will be discussed.

S12

Protein-protein Interaction Network for the Mammalian PP2A-type Phosphatases

A. Gingras, B. Raught, P. Shannon, A. Sanchez, R. Aebersold; Institute for Systems Biology, Seattle, WA, United States.

Protein phosphatases of the PP2A family (including PP2A and the related phosphatases PP4 and PP6) are widely expressed serine/threonine phosphatases, involved in a broad array of cellular processes. However, the regulation of this family of phosphatases remains poorly understood. We thus initiated a project aimed at identifying the protein-protein interaction network surrounding each of these important molecules, using a combination of affinity purification and mass spectrometry. Our approach involved the stable expression in human 293 cells of recombinant fusion proteins tagged at their amino or carboxy-termini with dual affinity tags separated by a specific protease cleavage site. The tagged proteins were subjected to the tandem-affinity purification (TAP) approach, thereby recovering the tagged protein (bait) and its associated binding partners (prey). Purified protein complexes were then directly cleaved with trypsin, and analyzed by LC-MS/MS. Our approach was iterative: when a novel interacting protein was identified, it was in turn cloned, expressed, and purified using the generic TAP-tag purification scheme. This strategy has proven highly effective. Our PP2A interaction network in mammals now consists of more than 50 individual proteins (including 15 previously uncharacterized molecules), establishing over 150 interactions. The function of some of the novel proteins will be discussed.

S13

A Map of the Interactome Network of the Metazoan *C. elegans*

S. Li, M. Vidal; Dana-Farber Cancer Institute, Boston, MA, United States.

To initiate studies on how protein-protein interaction (or “interactome”) networks relate to multicellular functions, we have mapped a large fraction of the *Caenorhabditis elegans* interactome network. Starting with a subset of metazoan-specific proteins, more than 4,000 interactions were identified from high-throughput yeast two-hybrid (Y2H) screens. Independent co-affinity purification assays experimentally validated the overall quality of this Y2H dataset. Together with already described Y2H interactions and interologs predicted *in silico*, the current version of the Worm Interactome (WI5) map contains ~5,500 interactions. Topological and biological features of this interactome network as well as its integration with phenome and transcriptome datasets lead to numerous biological hypotheses.

S14

Mapping Transcription Regulatory Networks in the Nematode *C. elegans*

M. Walhout; UMass Medical School, PGFE, Worcester, MA, United States.

Since the advent of functional genomic methodologies such as micro-arrays, vast amounts of gene expression data have become available. However, the transcription regulatory mechanisms responsible for the gene expression profiles observed frequently remain elusive. This is mainly because the identity of the regulatory transcription factors (TFs) that control differential gene expression often cannot be inferred from such datasets.

Regulatory TFs activate or repress transcription of their target

genes by binding to *cis*-regulatory elements that are frequently located in or near a gene’s promoter. Thus, in order to experimentally “match” a regulatory TF with its target genes, assays are needed that demonstrate a physical interaction between regulatory TFs and the *cis*-regulatory elements they bind to.

To facilitate the identification of metazoan TF-promoter interactions, we developed a Gateway-compatible yeast one-hybrid system that enables the rapid identification of TF-DNA interactions using both small (i.e. repeats of DNA elements of interest) and large DNA fragments (i.e. gene promoters). The possibility to generate many yeast one-hybrid DNA bait constructs simultaneously by Gateway cloning makes this system amenable to high-throughput settings. Here, we present protein-DNA interaction data obtained using various DNA fragments as baits including *Caenorhabditis elegans* promoters. Because of the ability to associate metazoan TFs with their respective target genes, the high-throughput Y1H system may prove to be a valuable tool to decipher global transcription regulatory networks that control differential gene expression in multicellular organisms.

S15

Standardized quantitative RT-PCR (StaRT-PCR), a sensitive means of assessing micro-array genome analyses of normal and transformed human oral keratinocyte lines

R. C. Grafström, Sr., M. Vondracek, J. A. Nilsson, J. Hoog, C. A. Staab; Karolinska Institute, Stockholm, Sweden.

Our laboratory is engaged in the development of highly defined culture models that mimic human oral epithelium, including exploration of the process underlying malignant transformation. Normal (NOK), SV40 T antigen-immortalized (SVpgC2a) and malignant (SqCC/Y1) oral keratinocytes have been cultured and studied in both monolayer and organotypic states. Transcript profiling using several types of Affymetrix microarray chips indicated expression of 4-5 x 10³ genes in the cell lines. The changes from *in vitro*-immortalization (SVpgC2a versus NOK) outnumbered those associated to malignant transformation (SqCC/Y1 versus NOK). Decreased expression/loss of genes was more common than increased expression/gain of genes. The noted alterations related to the cyto-skeleton, cell adhesion, differentiation, and oncogenesis. Other changes included transcripts related to xenobiotic-metabolizing cytochrome P450 enzymes, conjugation enzymes, and enzymes involved in detoxification of aldehydes and reactive oxygen. Alterations that infer retinoid insensitivity were also implied. The microarray results could mostly, but not entirely, be confirmed by other methods, including those for mRNA (Northern, *in situ* hybridization and StaRT-PCR) and protein (immunochemistry, 2D gel electrophoresis and enzyme activity) production. In this regard, the application of StaRT-PCR provided numerical data on transcripts to the level of 1-10 molecules per 10⁶ beta-actin molecules. Overall, the results indicated that several culture protocols might be used to mimic the phenotype of normal epithelial tissue with NOK cultures. In contrast, SVpgC2a and SqCC/Y1 mimicked severe epithelial dysplasia and well-differentiated squamous cell carcinoma, respectively. Clearly, application of genomics and proteomics methods have aided attempts to model normal tissue and the development of malignancy with cell lines, including mechanistic studies.

S16

Antibody microarrays and two-color rolling-circle amplification for sensitive, multiplexed serum protein measurements

B. B. Haab¹, H. Zhou¹, K. Bouwman¹, D. Dillon², J. Costa², P. Lizardi², R. Brand³; ¹Van Andel Research Institute, Grand Rapids, MI, United States, ²Yale University School of Medicine, New Haven, CT, United States, ³Evanston Northwestern Healthcare, Evanston, IL, United States.

Antibody microarrays enable highly multiplexed and rapid protein measurements in low sample volumes. The profiling of proteins in sera and other bodily fluids using this tool should offer new opportunities for biomarker discovery and insights into disease biology. We are using this tool to study the associations of many circulating tumor and host response markers with pancreatic and prostate cancers. Serum proteins that had been coupled to either a fluorescent tag (e.g. Cy3) or a hapten (e.g. biotin) were incubated on the microarrays, and specific proteins bound to the immobilized molecules on the microarrays through specific interactions. After washing away unbound proteins, bound proteins were detected using the fluorescent tag or amplified signal (using rolling circle amplification, RCA) from the hapten-labeled proteins. RCA significantly enhanced detection sensitivity while maintaining the accuracy and precision of the measurements. Multiple differences in serum protein expression patterns were observed between pancreatic cancer patients and healthy controls, including significantly higher levels of inflammation-related proteins and lower levels of negative acute-phase reactants in the disease samples. Several proteins were specifically elevated in pancreatitis patients, suggesting the possibility of using these measurements to specifically distinguish pancreatic adenocarcinoma from both healthy controls and patients with benign disease. Multi-parametric classification of the samples based on multiple proteins confirmed the concept of improved diagnostic accuracy through the use of combinations of markers. The analysis of many markers in parallel using antibody microarrays promises to have great value for cancer serum diagnostics.

S17

A Proteomic Approach to Autoimmune Disease

W. Robinson; Stanford University, Palo Alto, CA, United States.

The diversity of autoimmune responses poses a formidable challenge to development of antigen-specific tolerizing therapy. We developed 'myelin proteome' microarrays to profile the evolution of autoantibody responses in experimental autoimmune encephalomyelitis (EAE), a model for multiple sclerosis (MS). Increased diversity of autoantibody responses in acute EAE predicted a more severe clinical course. Chronic EAE was associated with previously undescribed extensive intra- and inter-molecular epitope spreading of autoreactive B cell responses. Array analysis of autoantigens targeted in acute EAE was used to guide the choice of autoantigen cDNAs to incorporate in expression plasmids to generate tolerizing vaccines. Tolerizing DNA vaccines encoding a greater number of array-determined myelin targets proved superior in treating established EAE, and reduced epitope spreading of autoreactive B cell responses. Proteomic monitoring of autoantibody responses provides a powerful approach to monitor autoimmune disease, and to develop and tailor disease- and patient-specific tolerizing DNA vaccines.

S18

Applications of Functional Proteome-Scale Microarrays

G. A. Michaud; Protometrix, Inc., Branford, CT, United States.

Comprehensive proteome-scale protein microarrays can be used to simultaneously screen up to several thousand proteins for drug binding, molecular interactions, or enzymatic activity. We have used arrays containing the majority of the yeast proteome to reveal specific and novel protein-protein interactions, as well as interactions between proteins on the array and lipids, nucleic acids, and small molecules. In addition to yeast proteins, we have expressed, purified, and arrayed thousands of human proteins. Examples will be given about how arrays of human proteins have been used to screen for enzyme substrates, measure enzyme inhibition, and identify protein targets of drugs. In addition, we have demonstrated the utility of these arrays for determining the specificity of antibodies, thus providing a new tool for the development and characterization of therapeutic and diagnostic antibodies.

S19

Single Molecule Approaches to Genome Analysis

D. Schwartz; University of Wisconsin-Madison, Madison, WI, United States.

Our laboratory developed the Optical Mapping System to rapidly scan populations at the whole genome level. The fully automated, working platform uses microfluidic chips to uniformly deposit very large genomic DNA molecules on to a biochemically friendly surface modality that enables a broad spectrum of biochemistries for high density marker placement and analysis of genomic DNA directly extracted from cells. As such, this fully functional system creates single molecule nanoarrays, that allow the entire representation of a genome to be analyzed in ways that go beyond chip-based system where many indels, and rearrangement events cannot be detected; these aberrations are likely to produce observable phenotypes, often correlated with disease. Recently, "Shotgun" Optical Mapping was used to analyze the entire of *Yersinia pestis*, *Shigella flexneri*, *Leishmania major*, *Escherichia coli* O157:H7, *Rhodobacter sphaeroides*, *Deinococcus radiodurans*, and *Plasmodium falciparum* (the major causative agent of malarial disease), and *Homo sapiens*, without the use of PCR, electrophoresis, or clones. Presently we are applying Shotgun Optical Mapping to the analysis of populations, including human, as well as of numerous microorganisms, where our mapping efforts are offering new routes to understanding genome variation and plasticity across individuals and closely related species. These efforts are also helping to facilitate the ongoing microbial sequencing projects at JGI, in terms of providing means for validation and aids for assembly. With the advent of a high-throughput Optical Mapping System, we are developing novel approaches for human association studies using a new class of genome markers that are designed to encompass SNPs (Single Nucleotide Polymorphisms), yet reveal genome variation on a scale not previously discerned for large populations. Current thinking in the field is centered on the use of a limited number of SNPs to leverage the apparent state of linkage disequilibrium, which is indicative of a young species; however, current approaches based on chips or mass spectrometry are pendant on huge numbers of oligonucleotides. This requirement limits analysis to a series of discrete loci and renders such approaches inadequate for the assessment of a broad spectrum of genome variation motifs. This limitation of current systems used for large-scale association studies may neglect discovery of important factors contributing to complex

traits. In this regard, haplotyping is emerging as the means to perform detailed analysis of mutations and is expected to play a major role in the field of pharmacogenomics. The Optical Mapping platform is uniquely suited for haplotyping using a broad range of marker types since analysis of single molecules allows for the unambiguous phasing of genetic markers within populations of unrelated individuals.

S20

Neurogenome Makes Neurotranscriptome Makes Neuroproteome

D. J. Smith; UCLA School of Medicine, Los Angeles, CA, United States.

The discussion will be concerned with new methods we have developed for genome-scale mapping of brain gene expression patterns in both health and disease. These technologies combine the mathematical concepts of biomedical imaging with the high-throughput tools of genomics. One approach, voxelation, employs analysis from the brain of spatially registered cubes or voxels, while another approach, gene expression tomography, employs analysis of axially rotated sets of slices. In both methods, image reconstruction results in volumetric maps of gene expression analogous to those obtained from biomedical imaging systems. The ultimate goal of our research is to create maps for all relevant biomolecular domains, including the transcriptome and proteome, allowing us to decode the logic by which the wiring diagram of the genome gives rise to the neuronal wiring diagram of the brain.

S21

Comprehensive, Quantitative and High Throughput Proteomics

D. G. Camp, II; Pacific Northwest National Laboratory, Richland, WA, United States.

The current explosion in proteomics technologies, largely due to the successes of genome sequencing efforts, have laid the foundation for new and higher throughput protein measurements as well as raised awareness of the limitations associated with mRNA expression measurements. Increasing evidence indicates that the correlation between gene expression and protein abundances can be low and that the correlation between gene expression and gene function can be even lower. Our substantially incomplete understanding of biological processes and incomplete knowledge of biomolecular players in pathways related to the progression of specific disease states contribute to the desire for comprehensive proteomic measurements.

The challenges associated with making useful comprehensive proteomics measurements include identifying and quantitating large sets of proteins whose relative abundances can span more than six orders of magnitude, vary broadly in chemical and physical properties, have transient and low levels of modifications, are subject to endogenous proteolytic processing, and for which individual proteins may be present at very low copy number per cell. The analytical needs to address these challenges far exceed what can be realized with existing technologies based upon tandem mass spectrometry.

Technology being developed at Pacific Northwest National Laboratory promises to provide more comprehensive, quantitative, and high throughput proteomic measurements. Analytical approaches for characterizing either intact proteins or global protein digests make use of accurate peptide mass and time tags. Attractions of these approaches include automated high-confidence protein identification, broad and unbiased

proteome coverage, and the ability to exploit stable-isotope labeling methods for high precision, relative protein abundance measurement. Results from fundamental studies of microbial systems and efforts aimed at developing biomarkers in mammalian systems for a variety of purposes will be presented to illustrate the approaches.

S22

Computational tools for statistical validation of high throughput proteomics data.

J. Eng, A. Keller, A. Nesvizhskii, R. Aebersold; Institute for Systems Biology, Seattle, WA, United States.

As increasingly large datasets are being generated from tandem mass spectrometry proteomics experiments, the need for statistically robust software tools that assign accurate probabilities to peptide and protein assignments is required. We will describe our progress in developing the methods and implementation of such tools, specifically PeptideProphet and ProteinProphet, for the automated data analysis of tandem mass spectrometry database search results. Additional information on the tools described in this presentation, including links to download programs and their source codes, can be found at www.proteomecenter.org/software.php.

S23

MS/MS - The Perfect Match

E. A. Kapp¹, F. Schütz², T. Speed², R. Simpson¹; ¹Ludwig Institute For Cancer Research, Melbourne, Australia, ²The Walter and Eliza Hall Institute of Medical Research, Melbourne, Australia.

The technology that has enabled proteomics to become a reality is tandem mass spectrometry (MS/MS), in which peptides derived from cellular proteins are ionised and fragmented, and the mass of each fragment is measured. Because peptides fragment inside a mass spectrometer in a relatively predictable way, the measured masses can be compared to the masses that would be predicted from the fragmentation of every peptide from every protein derived from every gene in the human genome database. Peptides are then identified by finding the best match between the measured masses from MS/MS spectra and the predicted masses from the human genome. Current search algorithms that match measured masses and predicted masses largely ignore information such as the peak intensities which can be a source of false positive matches in proteomic data. We have surveyed the MS/MS spectra of over 5000 peptides and found that fragmentation patterns are dependent on peptide ion charge state as well as amino acid composition in an ion trap mass spectrometer, which has led to the derivation of a new classification scheme termed the "relative proton mobility scale". Fragmentation models that incorporate this scale have been developed and tested to determine whether measured MS/MS spectra can be more accurately matched with predicted peptide fragmentation patterns, which should in turn increase the accuracy of protein identifications using tandem mass spectrometry.

S24

Reducing False Positive MS/MS Interpretations through Evaluation of the Quality of a Particular Spectrum in Relation to its Highest Scoring Candidate Peptide in a Database Search

K. R. Clauser; Millennium Pharmaceuticals Inc., Cambridge, MA, United States.

A typical proteomics LC/MS/MS experiment yields a collection of MS/MS spectra with wide variation in the extent of peptide fragmentation. This leads to a similarly wide distribution in the scores resulting from interpretation of those spectra from database search programs. A particularly troublesome source of false positive interpretations arises from the difficulty in segregating mediocre quality spectra with medium scoring, correct interpretations from high quality spectra with medium scoring, incorrect interpretations resulting from the effective absence of the correct answer in the database due to such factors as unexpected chemical modification, non-specific enzymatic cleavage, incorrect/ambiguous parent charge assignment. Approaches will be described for reducing false-positives through calculation of expected scores of both correct and incorrect interpretations for each individual spectrum as opposed to average expected scores for collections of spectra.

S25

The Chromatome: Mapping Protein:Chromatin Interactions in Mammalian Cells

L. Penn; Ontario Cancer Institute, Toronto, ON, Canada.

The development and application of ChIP-on-chip in mammalian cells enables the network of protein:chromatin interactions that occur *in vivo* to be profiled for the first time. We have used this technology to explore both the regulation and function of the Myc oncoprotein as a regulator of gene transcription during tumorigenesis. To facilitate the application of this technology on a global level, the Toronto Microarray Center is producing microarrays spotted with mammalian regulatory region DNA and will be distributing both human and mouse arrays in the near future. By this approach, a common platform can be used by a broad range of researchers to explore protein:DNA interactions in normal and diseased cells. Data integration will enable the chromatome to be visualized and new insights and opportunities for disease diagnosis and treatment to be realized.

S26

Using Chromatin Immunoprecipitation and Genomic Microarrays to Identify *In Vivo* Binding Site for Mammalian Transcription Factors

P. J. Farnham, M. Oberley, A. Kirmizis, D. Inman; University of Wisconsin, Madison, WI, United States.

We have combined the chromatin immunoprecipitation (ChIP) assay with the use of genomic microarrays to allow an unbiased identification of target genes of human transcription factors. A main advantage to our approach is that only DNA sequences directly bound by a factor within the context of a living cell are identified. Therefore, artifacts associated with overexpression or *in vitro* assays are avoided. By using immunoprecipitated chromatin to probe genomic microarrays that contain thousands of clones for different human promoters, we have identified novel binding sites for E2F and retinoblastoma family members. Our initial experiments were performed using human CpG island microarrays. We have now developed similar mouse CpG island microarrays that contain more than 5000 unique clones. Our studies using the human and mouse CpG arrays to identify binding sites for E2F family members and to study changes in histone modifications mediated by these factors will be presented. Because more than 60% of mammalian genes have CpG islands at their 5' ends, the ChIP-CpG assay provides a high-throughput approach to identify target genes of many mammalian transcriptional regulators. However, not all promoters are represented on CpG

arrays. Therefore, we are also developing additional arrays that will allow the analysis of all known human and mouse genes. Our progress towards the development and use of these new arrays will also be presented.

S27

Discovery of Transcription Factor Binding Sites Using High-Density Oligonucleotide Microarrays

R. D. Green¹, M. Oberley², J. Norton¹, M. Singer¹, P. Farnham²; ¹NimbleGen Systems, Madison, WI, United States, ²University of Wisconsin, Madison, WI, United States.

We present results demonstrating the utility of using custom oligonucleotide microarrays for identifying novel binding sites for transcription factors in the human genome. We designed several custom oligonucleotide arrays including an array containing 390,000 60mer probes that tiled through 30Mb of sequence for the Encyclopedia of DNA Elements (ENCODE) project. We hybridized labeled chromatin immunoprecipitation (ChIP) fragments generated with antibodies to E2F4, E2F6 and Pol II to these arrays and identified putative binding sites by looking for regions enriched in the ChIP samples compared to a total genomic and an IgG control sample. Putative binding sites were confirmed with PCR. These results demonstrate the utility of using custom oligonucleotide arrays for identifying transcription factor binding sites in any region of interest in a genome.

TUTORIAL SESSION SPEAKER ABSTRACTS

T1

Multicolor Molecular Beacons for Diagnostic PCR Assays

F. R. Kramer; The Public Health Research Institute, Newark, NJ, United States.

Molecular beacons are hairpin-shaped nucleic acid probes that undergo a fluorogenic conformational change upon binding to their target. As a consequence of their unique structure, they are extraordinarily specific. A single nucleotide mismatch between the probe sequence and the target sequence prevents hybridization. Molecular beacons can be labeled with differently colored fluorophores, enabling multiplex, homogeneous amplification assays to be carried out in real-time in sealed reaction tubes. Applications include the detection of infectious agents in clinical samples, the quantitative assessment of gene expression, the determination of genotypes, the identification of single-nucleotide polymorphisms, pharmacogenomic screening, and the determination of drug susceptibility.

T2

Pitfalls of Quantitative Real-Time PCR

S. Bustin; University of London, London, United Kingdom.

Real-time PCR assays are increasingly used for the quantification of mRNA levels not just from comparatively undemanding sources such as tissue culture cells, but from less-than optimal tissue samples such as *in vivo* biopsies, body fluids as well as archival and post-mortem samples. Since the result of any real-time PCR assay is critically reliant on the quality of the starting template, this creates uncertainty with respect to the validity, relevance and comparability of any data obtained from such sources. Furthermore, the reliability of quantitative real-time PCR results also depends on the quality of the reagents, operators and methods of analysis employed. This presentation aims to highlight the steps required to obtain

reproducible as well as biologically relevant data from *in vivo* and archival samples.

T3

The Multidimensional Protein Identification Technology (MudPIT): Variations and Applications

J. D. Venable, J. R. Yates III; The Scripps Research Institute, La Jolla, CA, United States.

In recent years, advances in a variety of technologies have provided the framework necessary to pursue large-scale analyses of proteins in complex mixtures. Because of the inherent need for speed and sensitivity in these analyses, mass spectrometry based strategies have excelled. One strategy that has been used extensively is named the multidimensional protein identification technology (MudPIT). In this approach, large numbers of peptides are separated by multi-dimensional liquid chromatography and subsequently identified by tandem mass spectrometry (MS/MS) coupled with database searching. MudPIT provides many advantages over traditional protein identification strategies including speed, sensitivity, and robustness. Moreover, because proteins are not subjected to gel electrophoresis before analysis and all the chromatography is performed on-line, the technique is easily automatable, and capable of detecting low abundance proteins. MudPIT is also a flexible strategy that has been applied to a variety of different areas of interest in proteomics including the identification of proteins in complex mixtures, the identification and mapping of modification sites, and quantitative studies using stable isotopic labeling strategies. We will discuss technical issues surrounding the use of MudPIT as well as variations of the technology that can be useful in some of these different applications.

T4

Comprehensive Analysis of Mammalian Cells and Serum Proteomes Using 3-D, 4-D, and 5-D Protein Profiling Strategies

D. W. Speicher; The Wistar Institute, Philadelphia, PA, United States.

Current protein profiling methods, including common top-down and bottom-up approaches, are capable of routinely comparing only about 1,000 to 2,000 proteins. This extent of protein coverage may be adequate for many prokaryote and yeast projects, but it represents a very minor portion of the proteins present in complex proteomes such as mammalian cells, tissues, and biological fluids. The best strategies for substantially increasing proteome coverage in quantitative profile comparisons involve inserting addition separation modes to achieve efficient 3-D, 4-D, and even 5-D methods. One very effective early stage separation is to subdivide complex proteomes into a modest number of well resolved fractions using solution IEF. A solution IEF prefractionation method that is particularly well scaled for proteome analysis is MicroSol-IEF using the ZOOM IEF Fractionator. The resulting pools can be analyzed using one of several alternative down stream profiling methods. For example, ZOOM IEF can be combined with slightly overlapping narrow range 2-D gels and large pore 1-D gels to profile more than 8,000 protein spots for cell extracts. Profiling of biological fluids such as human serum or plasma can be further enhanced by depleting the most abundant proteins using polyclonal antibody affinity chromatography followed by ZOOM IEF and 2-D gels. A particularly promising alternative strategy for serum as well as cell extracts utilizes a novel 3-D "Batch Protein Array-

Pixelation" method that can be readily coupled to additional separation dimensions. For example, a 4-D method for serum analysis utilizing Protein Array-Pixelation is: 1) affinity depletion of six high abundance proteins, 2) ZOOM IEF fractionation, 3) 1-D SDS PAGE of all fractions followed by dividing each lane into a uniform number of slices or pixels and digesting them with trypsin, and 4) nanoLC-MS/MS analysis of the trypsin digests from each pixel. Two 5-D variants of this method involve: 1) inserting several salt cuts on a SCX resin between the SDS gel and the reverse phase separation or 2) using an affinity tag such as cI-CAT to simplify the tryptic digests prior to LC-MS/MS. Hence, these 5-D Protein Array-Pixelation methods include 3 protein separation modes and 2 peptide separation modes that are largely orthogonal and resolution achieved at each step is well preserved during transfer to the next step. These methods can provide more extensive proteome coverage than most alternative profiling methods for complex proteomes. Although throughput is currently low, all steps can be automated and it is very likely that analysis times can be substantially reduced through further method optimization/streamlining and future instrumentation advances.

T5

Aptamers to Modulate Protein Activity

M. Stanton; Archemix Corp., Cambridge, MA, United States.

In the simplest view, aptamers can be thought of as nucleic acid analogs to antibodies - they are able to bind specifically to proteins, and, in many cases, that binding leads to a modulation of the protein activity. New aptamers are rapidly generated through the SELEX process and have very high affinity and specificity (pM to nM). Furthermore, aptamers composed of modified nucleotides have a long biologic half-life (hours to days), are non-toxic and non-immunogenic, and are easily produced using standard nucleic acid synthesis methods. These properties make aptamers ideal for both target validation and as a new class of therapeutics.

As target validation tools, aptamers provide important information that is complementary to that provided by other methods. For example, RNAi is widely used to demonstrate that protein knock-out in a cellular assay can lead to a biological effect. Aptamers extend that information by showing that dose-dependant modulation of the protein activity can be used to derive a therapeutic benefit. That is, aptamers can be used to demonstrate that the protein is a drugable target. As a new class of therapeutics, aptamers bridge the gap between small molecules and biologics. Like biologics, biologically-active aptamers are rapidly discovered, have no class-specific toxicity, and are adept at disrupting protein-protein interaction. Like small molecules, aptamers can be rationally engineered and optimized, are non-immunogenic, and are produced by moderate cost, scalable chemical procedures. As such, aptamers are emerging as an important source of new therapeutic molecules.

T6

Aptamers as Measuring Reagents

L. Gold; SomaLogic, Boulder, CO, United States.

Aptamers have been selected against a large variety of target molecules, from simple sugars and amino acids to large proteins and even whole viruses. In general aptamers have acceptable affinity for most research applications, and have in addition remarkable specificity. The known structures of several aptamers with their cognate bound analytes rationalize aptamer binding specificities.

SomaLogic has focused its efforts on creating large arrays of a special class of aptamers, such arrays to be used to measure precise concentrations of a variety of proteins (or other molecules) in different biological matrices (serum, urine, tissue extracts, etc). The special class of aptamers are called photoaptamers. These aptamers are single-stranded nucleic acids (the classic aptamers) containing a photo-activatable nucleotide that allows covalent binding to target proteins. The covalent binding increases specificity and enhances the analytical power of the arrays.

The first arrays have now been used in early clinical studies, seeking to identify protein signatures for various diseases. The data, coupled with simple multi-dimensional analysis, provide a reasonable hope that clinical diagnostics will become a precise means by which patient health status, disease onset, disease response to medical intervention, and disease recurrence will be followed.

T7 Quantifying Protein Interactions Using Surface Plasmon Resonance

Y. N. Abdiche; University of Utah, Salt Lake City, UT, United States.

Characterizing macromolecular interactions in detail is crucial to understanding their natural function. Surface plasmon resonance (SPR) is a powerful biophysical tool that can probe the affinity and kinetics of non-covalent complex formation between proteins and their diverse substrates. Examples will be drawn from protein binding interactions with other proteins, nucleic acids, small molecules, and lipids. Careful experimental design and the application of proper data processing methods are key to obtaining high quality biosensor data that can be interpreted kinetically, which will be the theme of this talk. The results of a study organized by the *Molecular Interactions Research Group* in 2002, which characterized an enzyme/inhibitor model system by SPR and two other complementary solution-based methods, will be discussed to validate the surface-based approach.

T8 Tutorial session: Protein arrays to study cancer and autoimmune disease

W. Robinson^{1,2}, **B. Haab**³; ¹Stanford University, Palo Alto, CA, United States, ²GRECC, Palo Alto VA, Palo Alto, CA, United States, ³The Van Andel Research Institute, Grand Rapids, MI, United States.

Protein array tutorial: We will discuss development and application of protein arrays to study cancer and autoimmunity. Topics of discussion will include protein array technology development and refinement, obtaining disease and control samples, study design, application of proteomics technologies to analyze clinical samples, and data analysis.

T9 Optimization and Troubleshooting of Protein Sequencers

R. Apfelzweig¹, **J. M. Neveu**²;
¹Applied Biosystems Inc., Foster City, CA, United States,
²Harvard University, Cambridge, MA, United States.

The field of chemical (Edman) protein sequencing is dominated by the instruments of Applied Biosystems, as they are the only vendor currently manufacturing and supporting high sensitivity chemical sequencers in North America. This tutorial will cover key points in optimization and troubleshooting for the current

models of ABI instruments, the Procise™ 49X-HT and 49X-cLC. Specific topics will include PTH-amino acid chromatography, interpretation of fluid sensor error data, and software/computer issues for the three platforms currently in use, the Macintosh with 610A data analysis, Procise PC 1.0 with Windows NT, and Procise PC 2.0 with Windows 2000 and SequencePro data analysis. Other instruments still in service (as detailed by recent ESRG studies) include older ABI sequencers (473A, 476A, 477A), and Hewlett Packard sequencers (1005A and 241NC); a brief overview of optimization and troubleshooting of some of these older, unsupported instruments will also be presented.

T10 Mapping Modification Sites on Proteins.

A. Krutchinsky, E. Chang, B. T. Chait; The Rockefeller University, New York, NY, United States.

Posttranslational modifications can change the function and the fate of proteins, and as a result, mediate many important cellular events. Knowledge about the nature of these modifications, when and where they occur on a protein, and how the abundance of modified proteins changes in the response to a particular cellular event helps us to understand how proteins govern cellular function. Here we present several examples of the mapping of posttranslational modifications on proteins using mass spectrometric techniques. Our approach is based on systematic study of posttranslational modifications of selected proteins that are hypothesized to bear particular modifications. Proteins of interest that were predicted to have in vivo posttranslational modifications were purified and digested with trypsin. The resulting tryptic mixtures were deposited on a compact disk (CD) MALDI sample stage [1], and mass spectra were obtained in an in-house modified MALDI-QqTOF (Centaur, Sciex) [2]. After analysis of the MALDI-MS spectra, the CD was transferred to an in-house constructed MALDI-ion trap mass spectrometer [1] for detailed MALDI-MS/MS analysis of the tryptic peptides that can bear modifications. We implemented MALDI-MS and MALDI-MS/MS to characterize phosphorylation sites on a group of proteins from *Saccharomyces cerevisiae*. Comprehensive computer analysis of yeast open reading frames produced a list of ~130 proteins that contain at least five cyclin dependent kinase (cdk) consensus motifs (S/TPXK/R). We have expressed and purified a number of proteins from this list and find that some of these proteins are indeed phosphorylated on several of the predicted sites. Comparison of these results with the results of in vitro kinase assay indicates that the analyzed proteins could cdk substrates. We also show that our mass spectrometric technique is suitable for characterizing other kinds of modifications. As an example, we show the change in methylation of histones H3 in wild vs. mutant mice in which a protein hypothesized to have a methyl transferase activity was conditionally knocked out.

Finally, we will discuss the applicability of this technique to the characterization of wider classes of modifications on proteins.
1. Krutchinsky et al. *Anal Chem.* 2001, 1;73(21), 5066-5077.
2. Krutchinsky et al. *J Am Soc Mass Spectrom.* 2000, 11(6), 493-504.

T11 Hypothesis-driven Proteomics

M. Kalkum¹, B. T. Chait²; ¹Beckman Research Institute of the City of Hope, Duarte, CA, United States, ²The Rockefeller University, New York, NY, United States.

Antibodies that combine high affinity with exclusive specificity are excellent tools to test for the presence of proteins of interest. However, high-quality immuno-reagents remain elusive or time-consuming to generate for a variety of targets. One alternative to antibodies involves mass spectrometric (MS) detection protein-specific proteolytic fragments. Unfortunately, MS observation of components from complex biological matrixes is often impaired by the limited dynamic range of the currently available methods. E.g., signal suppression effects and the so-called “chemical noise” can greatly reduce sensitivity and can often hamper reliable application of MS methods. We describe here a novel MS approach, termed *hypothesis-driven multistage mass spectrometry* (HMS-MS), that allows us to overcome these limitations. The detection principle is based on a search for the absence or presence of a predicted peptide in a manner analogous to the use of an antigen-specific antibody. Characteristic fragmentation patterns differentiate the signal from the background noise and provide reliable evidence for the presence of the predicted peptide. We present a broad range of applications of our approach. For example, we demonstrate a rapid (< 1 min) screen for determining mating types of yeast cells (as used in our lab for yeast genetic projects). Another example describes the detection of autoinducing peptides from supernatants of pathogenic Staphylococci that are involved in *quorum sensing* and *bacterial interference* and reveals several novel facts about their structures. The broad versatility of HMS-MS is also demonstrated for protein-interaction projects and for the analysis of posttranslational modifications.

T12

Data Tracking and Quality Control in High Throughput DNA Sequencing:

the ABRF DNA Sequencing Research Group (DSRG) 2004 Applied Technology and Informatics Tutorial

R. Pershad¹, R. Lyons², G. Grills³; ¹U.T. M.D. Anderson Cancer Center, Houston, TX, United States, ²University of Michigan, Ann Arbor, MI, United States, ³Harvard University, Cambridge, MA, United States.

As DNA sequencing laboratories significantly increase sequencing throughput, it is crucial to implement a comprehensive system for sample tracking, data distribution and quality performance evaluation. We will present how a mid-sized Core Facility and a high throughput Genome Center have developed and applied Laboratory Information Management Systems (LIMS) to support sample tracking, instrument performance, data distribution, financial management and laboratory administration. We will show how such information systems are used to generate quality control information on both sample processing and sequence production. Such information tracking allows facilities to respond quickly to sample problems, to troubleshoot failures and correct problems, to identify and correct subtle patterns of instrument and reagent failure, and to develop powerful statistical overviews regarding the strengths and weaknesses of a facility operation. Well designed information systems can help DNA sequencing laboratories achieve high levels of cost-effectiveness without compromising quality. Moreover, LIMS can be used as a tool to assist in maximizing sequence quality and read length in a high throughput laboratory. We will show practical implementations of information systems that provide quality control tracking to help monitor and troubleshoot production sequencing. We will present evaluations of new sequencing techniques, such as nanoliter scale cycle sequencing, and show how data tracking approaches have been

used to facilitate the development of new sequencing technologies.

T13

Microarray Based-Comparative Genomic Hybridization - Platforms and Applications

D. G. Albertson¹, J. R. Pollack²; ¹UC San Francisco, San Francisco, CA, United States, ²Stanford University School of Medicine, Stanford, CA, United States.

As originally described, comparative genomic hybridization (CGH) detects and maps DNA sequence copy number variation throughout the entire genome onto a cytogenetic map supplied by metaphase chromosomes. The use of metaphase chromosomes as the hybridization target has previously limited the resolution of CGH to 10-20 Mb, prohibited resolution of closely spaced aberrations, and only allowed linkage of CGH results to genomic information and resources with cytogenetic accuracy. Recently, metaphase chromosomes have been replaced with a variety of microarray platforms comprised of large insert bacterial artificial chromosomes (BACs), cDNAs or oligonucleotides. Relative copy number is measured at these specific loci by hybridization of fluorescently labeled test and reference DNAs as in conventional CGH. In array CGH, DNA copy number variations are measured across the genome with a resolution that depends primarily on the genomic spacing between clones. The technique is finding significant applications in cancer and medical genetics. This workshop will provide an overview of these platforms. Critical issues related to the use of the various array types for genomic DNA copy number measurements will be discussed, including the types and amounts of specimen required, protocols and data analysis procedures. Studies carried out on the arrays will be presented to illustrate the range of applications in medical genetics and cancer.

RESEARCH GROUP ABSTRACTS

R1-M

DNA Sequencing Research Group Presentation and Open Discussion: A Web Based DNA Sequencing Troubleshooting Resource

H. Escobar¹, J. W. Hawes², H. Lin³, D. S. Needleman⁴, T. Hunter⁵, E. Jackson-Machelski⁶, K. L. Knudtson⁷, R. Pershad⁸, M. Robertson⁹, G. J. Wiebe¹⁰; ¹Brigham Young University, Provo, UT, United States, ²Indiana University School of Medicine, Indianapolis, IN, United States, ³University of Utah, Salt Lake City, UT, United States, ⁴Thomas Jefferson University, Philadelphia, PA, United States, ⁵Vermont Cancer Center, Birmingham, VT, United States, ⁶Washington University School of Medicine, St. Louis, MO, United States, ⁷University of Iowa, Iowa City, IA, United States, ⁸The University of Texas M.D. Anderson Cancer Center, Houston, TX, United States, ⁹Ernest Gallo Clinic and Research Center, UCSF, San Francisco, CA, United States, ¹⁰Max Planck Institute of Molecular Cell Biology & Genetics, Dresden, Germany.

Despite recent advances in the automated DNA sequencing process, periodic problems associated with instrument platforms, purification, and reaction processes continue to exist. The goal of the 2004 DNA Sequencing Research Group (DSRG) is to establish a web-based troubleshooting resource. It would contain a user-friendly interface and a public database that allow researchers to search or input a variety of sequencing problem/solution data. The examples could range from “red

rain” on gel based platforms to the “waterfall effect” on capillary-based systems. We anticipate that a comprehensive troubleshooting guide will be useful to members of the DNA sequencing community at large, and that combined experience may eventually lead to the emergence of more robust protocols. Any comments, suggestions and sequencing data with detailed run conditions are welcome for the initial setup of the new website.

This Research Group presentation will be followed by an open discussion. This gathering will serve as a forum to promote and facilitate information exchange regarding DNA sequencing. This may present issues that could be included in the evolving troubleshooting resource. Solicitation of problems and solutions will also be pursued via the ABRF listserv prior to the meeting.

R2-T

Synthetic Peptides as Certified Analytical Standards

H. A. Remmer¹, N. Ambulos², L. F. Bonewald³, J. J. Dougherty⁴, E. Eisenstein⁵, E. Fowler⁶, J. Johnson⁷, A. Khatri⁸, M. O. Lively⁹, N. Ritter¹⁰, S. T. Weintraub¹¹; ¹The University of Michigan, Ann Arbor, MI, United States, ²University of Maryland, Baltimore, MD, United States, ³University of Missouri, Kansas City, MO, United States, ⁴Eli Lilly and Company, Indianapolis, IN, United States, ⁵Center for Advanced Research in Biotechnology, Rockville, MD, United States, ⁶Xcellerex, Marlborough, MA, United States, ⁷Texas A&M University, College Station, TX, United States, ⁸Massachusetts General Hospital, Charlestown, MA, United States, ⁹Wake Forest University, Winston-Salem, NC, United States, ¹⁰NMR Biotech, Germantown, MD, United States, ¹¹The University of Texas Health Science Center, San Antonio, TX, United States.

The goal of the Peptide Standards Project is the production, characterization, validation, and establishment of three synthetic peptides as registered and certified peptide reference standards. This project is conducted by the Peptide Standards Project Committee (PSPC) of the ABRF in collaboration with the National Institute of Standards and Technology (NIST).

The project entails the following:

(1) Large scale synthesis of the standard peptides in high (>98%) purity; (2) Tests for stability and shelf life of the peptides, (3) Packaging of the peptides in small quantities for distribution and analysis by ABRF member laboratories; (4) Use of the ABRF member analyses to yield the required Certificate of Analysis as specified by NIST.

The Peptide Standards Project was originally initiated by joint interactions between the Quality Compliance Committee (QCC) and the Peptide Synthesis Research Group (PSRG). It is the first ABRF project funded by NIST. Quality control data of the final peptides, information on packaging and the Certificate of Analysis as well as the status of this project will be presented.

R3-S

Proteomics Research Group Study 2004: Differentiation of Protein Isoforms

D. Arnott¹, M. A. Gawinowicz², W. S. Lane³, T. A. Neubert⁴, K. D. Speicher⁵, C. W. Turck⁶, K. A. West⁷, J. A. Kowalak⁸; ¹Genentech, Inc., South San Francisco, CA, United States, ²Columbia University, New York, NY, United States, ³Harvard University, Cambridge, MA, United States, ⁴New York University School of Medicine, New York City, NY, United States, ⁵The Wistar Institute, Philadelphia, PA, United States,

⁶Max Planck Institute of Psychiatry, Munich, Germany, ⁷Cleveland Clinic Foundation – Cole Eye Institute, Cleveland, OH, United States, ⁸NIH, Bethesda, MD, United States.

Accurate protein identification sometimes requires careful discrimination between closely related proteins. Protein isoforms and/or protein homologues may differ by as little as a single amino acid substitution or posttranslational modification, yet these differences can have profound effects on the structure and function of the proteins. The ABRF-PRG04 sample was designed to study a simple mixture of 3 closely related proteins, each at 3 picomoles. The Proteomics Research Group (PRG) sent the sample to interested laboratories in the form of intact proteins, and participating laboratories were asked to identify the proteins and report their results. The primary goal of the PRG04 Study is to give participating laboratories a chance to evaluate their capabilities and practices with regards to sample fractionation (1D- or 2D PAGE, HPLC, or none), protein digestion methods (in-solution, in-gel, enzyme choice), and approaches to protein identification (instrumentation, use of software and/or manual techniques to facilitate interpretation), as well as determination of amino acid or posttranslational modifications. Compilation of submitted data will allow a comparison of the strategies used and aid in optimization of these techniques.

R4-M

An Overview of Current Methods Available for the Analysis of N- and O-linked Carbohydrates

P. Azadi; Complex Carbohydrate Research Center, Athens, GA, United States.

There is a great deal of interest in the structures of carbohydrates attached to proteins because of their influence on the biological activity of the glycoproteins. Carbohydrates have been shown to be important in many cell processes such as leukocyte-endothelial cell adhesion, bacterial infections and immunological recognition of tumor cells. The discoveries at the forefront of research have motivated new methods to be developed to study the structure of these complex molecules. *N*-linked glycosylation is by far the most common type of carbohydrate in glycoproteins. They are found attached to Asn through a *N*-acetyl glucosamine (GlcNAc) and often contain 10-20 monosaccharides. *O*-glycans of the mammalian glycoproteins are attached to Ser or Thr through a *N*-acetyl galactosamine (GalNAc) and are usually smaller than *N*-glycans. Complete structural characterization of these *N* and *O*-linked carbohydrates is always a challenge mainly due to the fact that more than one glycan structure can be present at a particular amino-acid residue e.g. Asn, which contributes to what is called heterogeneity at that particular site. Thus, it is a demanding task to determine the actual site of each glycosylation as well as the structure of all oligosaccharides attached to the glycosylated amino acid residues. An overview of current techniques available for release, profiling, and sequencing of oligosaccharides, such as various derivatization methods in combination with GC-MS, HPLC, MALDI-MS, ESI-MS, and NMR will be discussed with advantages and limitations of each method.

R5-T

Nucleic Acids Research Group (NARG) Taqman®

Primer/Probe Design Study

G. L. Shipley¹, P. S. Adams², Y. Bao³, D. S. Grove⁴, B. P. Holloway⁵, A. T. Yeung⁶;

¹University of Texas Health Science Center, Houston, TX, United States, ²Trudeau Institute, Saranac Lake, NY, United States, ³University of Virginia School of Medicine, Charlottesville, VA, United States, ⁴Pennsylvania State University, University Park, PA, United States, ⁵Centers for Disease Control and Prevention, Atlanta, GA, United States, ⁶Fox Chase Cancer Center, Philadelphia, PA, United States.

The Nucleic Acids Research Group (NARG) designed an empirical study to define parameters required to make an optimal primer/probe set for a 5'-nuclease (Taqman®) real-time PCR assay. Assay design can be one of the major rate-limiting steps in rapidly acquiring data from genes of interest. Although a large number of pre-made assays can be purchased from commercial vendors, it is much more cost effective to design and purchase your own primer/probe sets. There are general guidelines available concerning assay design, however, exactly how important each of these parameters are has not been studied in an empirical manner. Further, there may be as yet unknown factors that should be taken into account during assay design. The purpose of this study was to give investigators an opportunity to design what they felt were an optimal primer/probe set for a common transcript and then have them tested empirically for effectiveness. Members of the NARG synthesized the primers/probe sets and tested them using a plasmid containing the mouse IFN γ cDNA insert as a template standard. Effectiveness of primer/probe design was judged by PCR efficiency (slope), deltaRn, y-intercept and Ct. The goal was to provide an opportunity for participants to sharpen their skills and/or learn some new ones while demonstrating the principles of Taqman® primer/probe design.

R6-S

Nucleic Acids Research Group (NARG) Real-time PCR Survey

P. S. Adams¹, Y. Bao², D. S. Grove³, B. P. Holloway⁴, G. L. Shipley⁵, A. T. Yeung⁶;

¹Trudeau Institute, Saranac Lake, NY, United States,

²University of Virginia School of Medicine, Charlottesville, VA, United States, ³Pennsylvania State University, University Park, PA, United States, ⁴Centers for Disease Control and Prevention, Atlanta, GA, United States, ⁵University of Texas Health Science Center, Houston, TX, United States, ⁶Fox Chase Cancer Center, Philadelphia, PA, United States.

The Nucleic Acids Research Group (NARG) of the Association of Biological Resource Facilities (ABRF) conducted an on-line survey designed to "take the pulse" of the real-time PCR international community. Respondents were asked questions concerning the type of facility, personnel, instrumentation, types of assays, templates, extraction methods, controls, analysis, etc. Additional questions for those that run core facilities that offer real-time PCR as a service were included. Participation was anonymous. Results provided insight into the way real-time PCR is being used in a variety of laboratory settings. Results will be freely available to all online.

R7-M

Amino Acid Analysis: Alive and Analyzing

D. Chin¹, W. Lane², A. Le³, R. Harris⁴, P. Hunziker⁵, J. Schaller⁶; ¹Chiron Corp., Emeryville, CA, United States, ²Harvard Microchemistry and Proteomics Analysis Facility, Cambridge, MA, United States, ³Mulan Laboratory, San Jose, CA, United States, ⁴Genentech Inc., San Francisco, CA, United

States, ⁵University of Zurich, Zurich, Switzerland, ⁶University of Bern, Bern, Switzerland.

The Amino Acid Analysis Research Group will host a Roundtable of six leaders in the field addressing the current and future technologies used in amino acid analysis. The invited panel members will present and compare post-column ninhydrin detection, pre-column PicoTag, AccuTag, FMOC/OPA and emerging MS techniques. Among the topics covered will be: 1. Limits of detection and quantitation for standard and non-standard amino acids. 2. Dynamic range of a typical analysis. 3. Accuracy and precision under normal conditions. 4. Labor intensity, run times and available automation. 5. Cost.

As a Roundtable, ample time will be given for audience participation to serve as a forum for discussing the future of amino acid analysis and its possible impact on proteomics and metabolomics. Panel: Steve A. Cohen, Waters Corporation; Dan Strydom, Nebraska Wesleyan University; K. Michael Gibson, Oregon Health & Science University; Peter Hunziker, University of Zurich; Mark Longster, Biochrom Ltd.; Richard Thoma, Monsanto.

R8-T

The MIRC '02 Study: Assembly State, Thermodynamic, and Kinetic Analysis of an Enzyme/Inhibitor Interaction

Y. N. Abdiche¹, D. G. Myszka¹, F. Arisaka², O. Byron³, E. Eisenstein⁴, P. Hensley⁵, J. A. Thomson⁶, C. R. Lombardo⁷, F. P. Schwarz⁸, W. Stafford⁹, M. L. Doyle¹⁰; ¹University of Utah, Salt Lake City, UT, United States, ²Tokyo Institute of Technology, Yokohama, Japan, ³University of Glasgow, Glasgow, United Kingdom, ⁴University of Maryland Biotechnology Institute, Rockville, MD, United States, ⁵Pfizer Inc., Groton, CT, United States, ⁶Pfizer Inc., La Jolla, CA, United States, ⁷Celgene-Signal Research Division, San Diego, CA, United States, ⁸National Institute of Standards and Technology, Rockville, MD, United States, ⁹Boston Biomedical Research Institute, Watertown, MA, United States, ¹⁰Bristol-Myers Squibb Pharma, Princeton, NJ, United States.

The *Molecular Interactions Research Group* (MIRC) sought to evaluate the capabilities of Core Facilities and other research laboratories to analyze a small molecule/protein interaction by implementing three biophysical technologies: analytical ultracentrifugation (AUC), isothermal titration calorimetry (ITC), and surface plasmon resonance (SPR). A model system composed of an enzyme/inhibitor pair was distributed to a large panel of AUC, ITC, and SPR operators. Participants were asked to measure one or more of the following: 1) the molecular mass, homogeneity, and assembly state of the enzyme by AUC, 2) the affinity and thermodynamics for complex formation by ITC, and 3) the affinity and kinetics of complex formation by SPR. AUC confirmed that the native enzyme was monomeric in free solution at concentrations relevant to the ITC and SPR portions of the study. Binding affinities obtained by ITC and SPR were within identical experimental error and the van't Hoff enthalpy change determined by SPR was indistinguishable from the ITC-derived value. A direct comparison of solution- and surface-based measurements confirmed that immobilizing the protein on a biosensor surface did not modify its binding activity.

R9-S

FARG Study 2004: Whodunit? Analysis and Comparison of Equipment and Protocols used for Fragment Analysis

C. Rosato; Oregon State University, Corvallis, OR, United States.

The Fragment Analysis Research Group (FARG) of the Association of Biomolecular Resource Facilities (ABRF) conducted a study to measure accuracy and precision among equipment and methodology used in fragment analysis applications. The purpose for surveying participants was to gather data from a wide representation of laboratories using various types of equipment for genotyping. Because size variability can lead to genotyping errors, it is critical to determine whether a method for standardizing samples within a project can be achieved when the samples are processed in different laboratories. Our study was designed to evaluate size variability within and between different platforms (equipment and processing methods), determine whether genotyping can be successful despite this variability, give participants a method for self-evaluation in a study presented as a hypothetical DNA forensics situation. Samples were distributed by mail, and data were collected using both an online survey and an .ftp server. Results from the analysis of the collected data are presented.

R10-M

An ABRF Snapshot: ABRF 2003 Survey

J. L. Bleibaum¹, C. Nicolet², G. Sarath³, J. Simpson⁴, S. Yadav⁵, M. Young⁶;

¹Roche Palo Alto, Palo Alto, CA, United States, ²University of Wisconsin, Madison, WI, United States, ³USDA-ARS, Lincoln, NE, United States, ⁴SAIC-Frederick, Frederick, MD, United States, ⁵Cleveland Clinic Foundation, Cleveland, OH, United States, ⁶City of Hope, Duarte, CA, United States.

Technologies in use in core facilities are in flux. As new technologies are adopted, old ones are used less frequently or abandoned altogether. The ABRF Survey Research Group designed the ABRF 2003 Survey to examine technologies currently in use in core facilities as well as sample frequency and sample turnover. Technologies surveyed were categorized into four groups: mass spectrometry services, protein and peptide services, DNA services, and bioinformatics services. Respondents were also asked to describe other characteristics of their facility, including type of core facility (academic, industry, etc.), staffing levels, and funding sources. Preliminary results show strong trends in usage of specific technologies. For facilities offering mass spectrometry services, MALDI-TOF MS is the most common technology used. DNA sequencing and fragment analysis is primarily being carried out on capillary instruments; some labs still use gels. N-terminal sequencing and protein digestion are offered much more frequently than peptide synthesis or amino acid analysis. The most common bioinformatics service offered is database searching. The majority of the respondents are working in academic institutions; most core facilities are small labs of 2 to 3 scientists with a ratio of Ph.D. level scientists to B.S./M.S. scientists of 1:2.5. Funding for new instruments and new hires is often difficult, however facility support in the future is generally seen as positive. The data collected from this survey show that core facilities are important contributors to scientific research and that new technologies are adopted and utilized by core labs.

ROUNDTABLE DISCUSSION SPEAKER ABSTRACTS

RT1

Proteomic Analysis of Cellular Signaling

F. M. White; Massachusetts Institute of Technology, Cambridge, MA, United States.

Protein phosphorylation is one of the most prevalent post-translational modifications in the cell, and regulates much of the cellular signaling. Analysis of protein phosphorylation events on a global scale, if done well, should enable scientists to decode cellular signal transduction. Global analysis of protein phosphorylation is a daunting challenge: the phosphoproteome, comprised of approximately 30% of mammalian proteins, is extremely complex. Additionally, protein phosphorylation is a dynamic event; as a result the phosphoproteome is constantly changing. Our mass spectrometry-based technology allows us to interrogate hundreds to thousands of protein phosphorylation events in a single analysis, identifying the site of phosphorylation on specific peptides. This technology also allows for generation of relative quantification information from one cell state to another, allowing us to track changes in the phosphoproteome through the cell cycle or on treatment with a drug. This information should allow us to decode the signaling pathways regulating cellular transformation, and may lead to a new class of biological markers for both disease and drug efficacy. Decryption of cellular signal transduction changes on drug treatment will reveal the mechanism of action of a variety of small molecule drugs, and will also increase our understanding of the toxic effects of various compounds.

RT2

Phosphoproteomics by ESI/Q-FTMS

N. L. Kelleher; UIUC, Urbana, IL, United States.

For detection and localization of modifications such as phosphorylation, mass spectrometry (MS) at the peptide and intact protein level have complementary advantages. A host of phospho-peptide detection technologies have been reported of late, including those using intact proteins (*i.e.*, the Top Down MS approach) and Electron Capture Dissociation (ECD) of phosphorylated species to enable precise localization of phosphorylation sites. While the ejection of phosphate (-98 Da) from phospho-Thr and phospho-Ser residues continues to be understood and ameliorated, a quadrupole-Fourier Transform hybrid (Q-FTMS) now allows for enhanced ion populations to be subjected to ECD. A new software and database environment (called *ProSight PTM*) now allows improved detection abilities for both putative PTMs and those detected previously using any type of measurement technology.

RT3

Fragment Analysis Software Analysis Programs: Problems, Insights and Helpful Hints

C. Rosato¹, R. Scholl², M. Miller³; ¹Oregon State University, Corvallis, OR, United States, ²University of Utah, Salt Lake City, UT, United States, ³National Cancer Institute, Bethesda, MD, United States.

Fragment Analysis is a technique in which DNA fragments (usually PCR amplicons) are electrophoretically separated and analyzed. Research projects may include genotyping within and between families, DNA fingerprinting, mutation detection, and genotyping a community of unidentified individuals. While the details of these data represent diverse research questions, their value is directly proportional to the precision of the data analysis. Accurate sizing of the DNA fragments must precede any downstream analyses.

This Round Table and open discussion session will present two

topics on analysis software designed to work with fragment data. The first topic describes *GeneMapper*. Intended as a replacement for the two analysis programs GeneScan and Genotyper, *GeneMapper* has some advantages and pitfalls that will be addressed by the speaker including many useful hints for using this software.

The second topic will cover a newly written software program to augment data collected on ABI sequencers. *Chromagna* affords some new ways of working with data. For Genescan files, one can: add, delete, split or fuse peaks; define or redefine standard peaks; and separate overlapping stutter patterns. One can add, delete or modify standard peak designations interactively. Genescan can usually read the resultant files for post-analysis. Five dye files can be read, displayed and printed. *Chromagna* also has DNA Sequencing applications as well. The author of the software will introduce and describe its useful applications.

We encourage audience participation on these and any other analysis software used for fragment analysis applications.

RT4

Chromagna: A Program to Supplement Analysis of Sequencing and Genescan Data

M. J. Miller; National Cancer Institute, Bethesda, MD, United States.

Table of Contents

- Chromagna: A Program to Supplement Analysis of Sequencing and Genescan Data

Chromagna is a Mac-based program (OS 8.6, 9, or 10 in classic mode) written for analysis of data generated by ABI sequence instruments such as the 377 or 3100. A windows version of the program is being developed. The program has a number of features that can aid in the analysis of sequence or Genescan data. (1) Chromagna can read and display either Sequence or Genescan files. (2) It can remove "spikes" from capillary data files. These are the blips caused by either micro-bubbles or dust in the POP polymer. (3) Chromagna can flatten the background to remove the "Waterfall effect," a frequent problem in data coming off capillary sequencers. (4) The program can print chromatograms in varied and flexible ways. For example: print only current window; print the entire chromatogram on one or more pages; define the number of panels per page; define start and stop points. (5) The different data can be exported in comma separated value (".csv") text format. Many text processors, spreadsheets, and graphic programs can read this format. (6) With Genescan files, one can: add, delete, split or fuse peaks. One can also separate overlapping stutter patterns using the method of Miller & Yuan (*Analyt Biochem*, 251, 50-6 (1997)). (7) One can add, delete or modify Genescan standard peak designations interactively. (8) Sequences can be edited in the usual way. (9) Five dye files can be read, displayed and written to. Finally, there is a reference manual (MS Word format) that details the use of the program.

RT5

A Comparison of Multiple Microarray Platforms for Gene Expression

B. Merriman; University of California at Los Angeles, Los Angeles, CA, United States.

There are currently a variety of microarray platforms available for doing whole genome gene expression measurement. A new

Microarray Consortium sponsored by the NINDS and NIMH has been chartered to provide expedient access to such technologies for sponsored researchers. As a first phase of this effort, Consortium members have undertaken a large scale comparison of the currently available technologies, in order to characterize the levels of repeatability, accuracy, sensitivity and agreement between the diverse platforms. In this report, we summarize our findings. The study is based upon assaying aliquats from the same four total RNA samples (samples derived from bulk Human Liver, Kidney, and Spleen, plus the Stratagene Human Reference pool) on diverse platforms, each sample done in triplicate, with the assays carried out at centers/facilities well-versed in the respective techniques. The platforms include whole genome pre-fabricated oligo arrays from Affymetrix, Agilent and Amersham, plus custom spotted arrays from whole genome oligo sets from Operon, ClonTech and MWG, plus 33,000 clone custom spotted cDNA arrays. In addition to testing for repeatability and consistency, independent validation of accuracy and sensitivity is provided by assaying the same samples via large scale quantitative RT-PCR targeted on the 100 genes most discordant between microarray platforms, as well as SAGE analysis (a form of direct molecular transcript counting). We summarize the many dimension in which these approaches agree and disagree, and discuss the implications for platform choices and appropriate interpretation of microarray data.

RT6

Cell Arrays for Chemical Genomics and Systems Biology

J. Huang; Los Angeles, CA, United States.

We will describe a new "cell array" technology which automates and miniaturizes chemical genomic screening, allowing an efficient and thorough analyses of any drug's effect on cellular processes at the single-gene level over the whole genome. The cell array platform should be universally applicable to phenotype-based, high-throughput investigations of gene function, microbial genomes, infectious disease and drug resistance. The new technology should also greatly facilitate the identification of drug targets and the study of the mechanisms of drug action on a global scale.

RT7

Identifying the composition of proteins in biological complexes using multidimensional chromatography coupled with tandem mass spectrometry

A. J. Link; Vanderbilt University Medical School, Nashville, TN, United States.

Most biological functions are carried out by a network of proteins interacting either in a signaling complex or pathway. One of the most important challenges facing investigators in the post genome era is to define the context in which a given protein functions. To meet this challenge and define the native contexts in which a given protein functions, we employ various approaches to isolation and purification of target proteins under non-denaturing conditions to identify interacting proteins. Multidimensional chromatography coupled with tandem mass spectrometry protein identification is being developed to rapidly define large multiprotein complexes. Statistical and bioinformatics methods are applied to define the protein interactions and identify putative posttranslational modifications. Relevant approaches to isolate protein complexes followed by a multidimensional mass spectrometry analysis will be presented. A new complex network of protein interactions is emerging from our analysis.

RT8

The Development and Application of MudPIT as a Quantitative Proteomic Method

M. Washburn; Stowers Institute for Medical Research, Kansas City, MO, United States.

The dynamic changes of a proteome or fractions of a proteome, i.e. organelles and protein complexes, can be analyzed via quantitative proteomic methods. In quantitative proteomics, the relative abundances of the same protein, but from two different samples and with two unique masses, can then be measured. Metabolic labeling strategies are one approach to this whereby the mass of proteins is modified by growing an organism, such as *S. cerevisiae*, in media that has “heavy” or “light” nutrients. MudPIT is a chromatography based proteomic technique where a complex peptide mixture is prepared from a sample and loaded directly onto a biphasic microcapillary column packed with reversed phase and strong cation exchange HPLC grade materials. Once the complex peptide mixture is loaded onto the biphasic microcapillary column, this column is placed directly in-line with a tandem mass spectrometer. MudPIT is capable of carrying out quantitative proteomics analyses. By combining global analysis of mRNA expression and global quantitative proteomic analyses one can gain novel insight into the biology of an organism and begin to discern the interplay between mRNA and protein expression levels of a given locus in an organism’s genome. During this session we will discuss the development of MudPIT as a quantitative proteomics tool and its application to the analysis of the mRNA and protein expression change in *S. cerevisiae* cultured in different growth medias.

RT9

Multidimensional Separations in High-Throughput Quantitative Proteomic Analysis by Mass Spectrometry

T. Griffin; University of Minnesota, Minneapolis, MN, United States.

This presentation will describe the current methodologies used in large-scale quantitative proteomic analysis using mass spectrometry. The components of this methodology include: 1) Labeling of protein and peptide mixtures using stable-isotopes; 2) Separation and fractionation of the peptides using multidimensional separation methods; 3) Automated mass spectrometric analysis and sequence database searching. This presentation will cover the practical use of these methodologies, the inherent advantages and limitations, and recent advances.

RT10

MS Analysis of Glycoproteins

R. Orlando; CCRC/UGA, Athens, GA, United States.

Glycosylation is one of the more common modifications found on proteins. These carbohydrate side chains have been shown to play diverse biological roles ranging from immune surveillance to cell-cell signaling to cell development and differentiation. Unfortunately, protein glycosylation is one of the more difficult modifications to characterize.

This presentation (and round table discussion) is intended to serve as a primer on the MS characterization of glycoproteins. We will describe some of the typical problems associated with the mass spectrometric analysis of glycoproteins, which are rarely discussed and more seldom published. These challenging areas are associated with items such as glycan heterogeneity,

poor glycan ionization efficiency, and the large hydrodynamic volumes of carbohydrates. These factors cause a myriad of problems that are not encountered with other types of protein modifications.

We will also describe the current procedures for characterizing glycoproteins, which utilizes a combination of MS and chemical/biochemical procedures. This procedure enables the identification of N-/O-linked glycosylation sites and structural elucidation of the attached glycans. Additionally, the limitations of these procedures and areas that need improvement will be discussed.

RT11

Tandem Mass Spectrometry for Glycoproteomics

J. Zaia; Boston University, Boston, MA, United States.

Although the proteomic analysis task becomes vastly more complex when post-translational modifications are considered, it is essential to undertake measurements of protein glycosylation. Glycosylation occurs to many, if not most, cell surface and extracellular matrix proteins and plays important roles in cell-cell and cell-matrix recognition events. Significantly, many pathogens recognize cell surface carbohydrates, and the understanding of these events is crucial for the development of drugs and therapeutics. Collisional-induced dissociation (CID) mass spectrometry has become the enabling technology of proteomics because of the availability of genomic databases and the well understood fragmentation behavior of peptides. No such databases exist for glycans, however, and their fragmentation behavior is more complex than that of peptides. One factor contributing to this complexity is that N- and O- linked glycans are branched rather than linear. Another factor is that glycosidic bonds involving certain monosaccharides residues fragment more readily than do those of others. Glycans are also inherently heterogeneous in structure, increasing the number of variants for a given gene product. It is possible to observe ions corresponding to glycopeptide using the MALDI and ESI conditions used in proteomics, with a few modifications. Using CID MS, a glycopeptide ion fragments to form very abundant ions from cleavage of the glycan and very low abundance ions from peptide bond fragmentation. As a result, information on the peptide backbone is lacking. Using multistage MS, however, the ion corresponding to the deglycosylated peptide can be subsequently isolated and fragmented, producing the desired information. The use of electron capture dissociation (ECD) on a Fourier Transform mass spectrometer provides particularly useful information on post-translationally modified peptides in that only the peptide backbone is cleaved, allowing the modified amino acids to be readily identified.

RT12

Roundtable on MALDI TOF/TOF

D. F. Barofsky¹, **S. Martin**², **A. La Rotta Angenendt**³; ¹Oregon State University, Corvallis, OR, United States, ²Discovery Proteomics & Research Center, Applied Biosystems, Framingham, MA, United States, ³MALDI and ESI-TOF Division, Bruker Daltonics, Inc., Billerica, MA, United States.

Each of the three panelists will lead off the roundtable with a different perspective on the application of TOF/TOF to the analysis of proteins. These short, introductory presentations are intended to stimulate an interchange of questions, comments, and discussion between the panelists and participants in the audience and between the participants themselves. Steve Martin will open the roundtable with remarks on the

integration of chemistry, MALDI TOF/TOF technology, and informatics to address biological questions. The integration of MALDI with tandem time-of-flight instrumentation has broadened the type of biological applications that may be addressed with MALDI based workflows. Results-dependent acquisition is one of the attributes that is unique to MALDI based MS/MS approaches. Once the sample is deposited on the MALDI plate, it is frozen in time. This enables workflows in which MS and MS/MS spectra may be iteratively acquired and compared to databases of information to determine if more spectra are required to refine and validate results of biological experiments. This instrument/informatics based approach has been coupled with up front chemistries that enable quantification of the proteins within the sample to produce an overall workflow for differential protein expression analysis. This overall workflow will be discussed with relevant biological applications.

Doug Barofsky will comment on the use of off-line HPLC MALDI TOF/TOF as a complement to on-line HPLC ESI MS/MS analysis of proteins. Based on the analysis of a large data set containing hundreds of peptides and thousands of individual amino acids, some of the currently held notions regarding the complementary nature the MALDI and ESI processes have been confirmed recently in his laboratory. In general, MALDI is better disposed toward the analysis of basic and aromatic species whereas ESI tends to favor the analysis of hydrophobic amino acids and peptides. These tendencies in ionization, which account in large part for the complementary nature of the peptides and proteins identified by the MALDI and ESI instruments, will be summarized with a viewpoint that MALDI tandem time-of-flight MS must be used in conjunction with ESI MS/MS in order to gain the most information out of a given sample in a proteomics study.

Aurelio La Rotta A. will focus his opening observations on advanced strategies, current limitations, and future perspectives for TOF/TOF based protein analysis. These will touch on two concepts for reaching a new level of sequence coverage and reliability in protein identification. The first approach would allow ESI MS and MS/MS to be combined with quasi on-line MALDI TOF/TOF analysis on the chromatographic time-scale. The second, approach, which is complementary to the first, involves top-down analysis of undigested proteins with in-source decay (ISD) MALDI TOF/TOF. Technical aspects of these two approaches, preliminary data that demonstrate performance and limitations will be presented along with some personal views on what may be expected from an instrumental point of view in the near future.

RT13

FARG RoundTable: Non-Sequencing Methods for Detecting Unknown Single Nucleotide Polymorphisms

D. A. Bintzler; University of Cincinnati DNA Core Facility, Cincinnati, OH, United States.

As a resource facility, our laboratory has processed a large number of samples submitted for sequencing in order to identify single base changes, insertions and deletions. Investigators analyze these results and compare them to a wild type sequence to determine a possible relationship between the base changes and a genetic disorder. Therefore, a common goal among researchers is to identify single base polymorphisms (SNPs) in specific genetic regions isolated from patients with symptoms of a genetic disorder compared to healthy individuals (wild type). Successful discovery of many SNPs has led to establishment of important relationships between single base mutations and disease. Automated sequencing, as a

method of SNP discovery in our resource laboratory, is limiting in the number of samples that can be processed and often costly. Therefore, we researched potential new methods that could be used to identify unknown SNPs without sequencing. Investigations have led to a potential protocol in which wild type sequences are annealed to sequences with possible SNPs. The presence of single SNP or multiple SNPs creates base mismatches between the wild type strand and the test strand. A nuclease, capable of cleaving single stranded DNA, might recognize the mismatch and cut the wild type-mutant combination into shorter fragments that could be separated. The results of our research will be presented for open discussion.

RT14

Non-Sequencing Methods for Detecting Unknown Single Nucleotide Polymorphisms

C. Rosato¹, **D. Bintzler**², **A. Yeung**³; ¹Oregon State University, Corvallis, OR, United States, ²University of Cincinnati, Cincinnati, OH, United States, ³Fox Chase Cancer Center, Philadelphia, PA, United States.

Identification of single nucleotide polymorphisms (SNPs) can reveal important information about the phenotype of an individual organism. For example, a single base change or an indel (insertion or deletion) in a gene's coding region can alter the function or structure of the encoded protein. This change could lead to deleterious effects that might threaten an organism's survival. Even though most SNPs appear to be located in non-coding regions, their identification is still useful as they can be used as genetic markers in population studies. The completion of the human genome project allows medical researchers to identify and characterize SNPs and to begin studying their effects on whole organism physiology. Automated DNA sequencers are often used in the discovery and analysis of SNPs, and these instruments are often already present in labs that carry out DNA sequencing and/or fragment analysis.

Other methods used to analyze SNPs include real time (TaqMan) PCR, dHPLC, Pyrosequencing, and mass spectrometry. However, these methods are generally limited to identification of SNPs that have previously been identified through sequencing. While sequencing remains the standard method for identifying unknown SNPs, dHPLC is also widely used for SNP discovery.

The purpose of this Roundtable is to conduct an open forum based on non-sequencing methods for identifying unknown SNPs. The discussion will begin with short presentations on enzymatic cleavage of base mismatches that might indicate the existence of a SNP. An open forum where we encourage discussion of other methods of SNP identification will follow two brief presentations.

RT15

CEL I mismatch-specific endonuclease

A. T. Yeung, **E. Nicolas**, **L. Hancock**; Fox Chase Cancer Center, Philadelphia, PA, United States.

We discovered a family of nucleases, exemplified by CEL I from celery, that have high specificity for DNA mismatch, including base-substitutions, insertions and deletions. My laboratory continues to study CEL I enzymology and to optimize the use of CEL I on different fragment analysis platforms. Some platforms work best with the single-strand mismatch-specific nicking activity of CEL I while other platforms can benefit from the mismatch-specific double-strand DNA truncation activity at a higher CEL I concentration. The

former is highly effective on the ABI 3100 and the Beckman CEQ8000 genetic analyzer while the latter can exploit the high sensitivity of the Agilent Bioanalyzer lab-on-a-chip system without using fluorescent primers or DNA denaturation. The advantages of each CEL I method will be discussed. Optimization of mutation detection on these platforms will be shown. We found that CEL I, because it uses a different

mechanism of mismatch detection than Sanger sequencing, is effective at detecting mutations that are missed by DNA sequencing methods as well as pinpointing the location of the mismatch in the template DNA.
<http://web-apps.fccc.edu/fccc/yeung/index.html>