Conference Summary

THE EMERGING USE OF GENOMICS AND PROTEOMICS IN ENDOCRINOLOGY*

G R Sridhar

The Endocrine Society of America organized a *Hot Topics* symposium on 'The emerging use of genomics and proteomics in endocrinology' at New Orleans on November 15th and 16th, 2002. It was a focused session 'to provide an introduction, overview and practical approach to genomics and proteomics.' Both clinicians and basic scientists were provided an 'overview of existing and emerging technologies for analysis of specific proteomics.' 118 participants attended the symposium from the Americas, Europe and Asia.

The programme started with a keynote address by Jeffrey M Friedman on leptin, obesity and the genomics of energy homeostasis. He focused on three aspects: why is it difficult to maintain weight loss, fatty liver of obesity, and why are some people thin while others are not? Leptin is a fat cell derived signal that reduces food intake by acting on the hypothalamus. Mice deficient in leptin are obese, and lose weight when the hormone is replaced. However except those with rare syndromes, most obese humans are not leptin deficient. On the contrary, many obese type 2 diabetes persons have elevated leptin levels, suggesting a state of insulin resistance. Rather, leptin may have clinical application in insulin resistance of lipodystrophy, hepatic steatosis and in HIV infection. Hepatic steatosis was shown to respond to leptin administration.

SCD-1 gene expression in ob/ob mice can be knocked off by deletion of exon 1-4 of SCD-1. In mice, SCD-1 altered energy expenditure and appeared to determine whether fat was stored or burned off. Friedman suggested that Niel's thrifty gene hypothesis could explain why the metabolic syndrome X was common in populations that have recently shifted from hunter-gatherer existence to one of adequate food availability. However, he added that subtle evolutionary processes may be active, with differential expression of pre-existing gene variants, rather than de novo fresh mutations. It is not yet known why some individuals are lean. It boils down conceptually to a balance between energy intake and expenditure.

David M Altschuler spoke on 'Human genome sequence variation and the genetics of common endocrine diseases.' He began by stating we have currently entered the genomic era. To say it is now the post-genomic era is like saying the post-computer era begun after 1977, when the first personal computer was introduced. What must happen after the human genome has been sequenced is to assign functions to the many uncharacterized genes, and to correlate disease traits to variation in the expression of genes and proteins. Ultimately a global view of biology is to be expected. Technology is focusing on distinguishing signal from noise. Recombinatric mapping in families would suffice for monogenic diseases. The more common polygenic disorders require large samples for testing. Characterizing human genome sequence variation is a natural corollary of sequencing the human genome.

Human heterozygosity, he said, is relatively limited and is largely attributable to common variations. During evolution deleterious variants are selected out and are therefore rare; neutral variants occur similarly to that in the general population; beneficial variants become common. The selective history of disease alleles is not yet known. To get comprehensive associations, shared ancestor blocks (haplotype blocks) may be studied.

Epidemiological associations are generally not reproducible because the sample is not large enough, given the modest genetic influence. Ultimately human genome sequence variations impact on understanding mechanisms of diseases and stratify at risk sub groups. The challenges lie in issues of privacy, group identity and how they make an impact on medical science.

Robert A Hegels, addressing 'The promise of genomics and human endocrine disease' focused on

---

** Conference Summary of Endocrine Society of America, Nov 15th & 16th 2002.*

Endocrine and Diabetes Centre, 15-12-16 Krishnanagar, Visakhapatnam 530 002, grsridhar@hotmail.com

familial partial lipodystrophy, in a Canadian kindred where LMNA disease mutations were identified by disease gene mapping. Clinical acumen was used as a mapping tool along with statistics and biology. He also touched on HNF1 alpha as a candidate gene for maturity onset diabetes of young. Links would be ultimately established among social, biological, biochemical and physiochemical aspects.

Jim Woodgett reviewed microarray technology in his lecture 'Gleaning insights into dysfunctional signal transduction pathways from global gene expression profiles.' Microarrays are glass or silicon slides etched usually from 500 to 20,000 genes. Using conventional DNA hybridization process, relative expression of these genes is simultaneously measured between samples. Laser activated fluorescence readers are employed to read data, Microarray technology is easier than proteomics, but is less direct. It promises a broad based system view of biological perturbations. He presented work on custom spotted cDNA microarray studies on human immune related gene expression profile. Among SAM mutations, false signals are identified, post hoc analysis done, followed by visualization and validation. Validation is a critical step in the whole process.

Unlike metabolic pathway, gene pathways, signaling proteins are regulated not by transcription, but by rapid posttranslational processes. Signaling is modulated by extracellular matrix, which rather than being merely permissive can programme different gene expression in response to the same stimulus.

Speaking on 'Using genomics to understand hormone action,' Paul Meltzer commented that in performing modern biology, machines are many and people few. Genomics is likely to revolutionize not only biomedical sciences, but also plant biology, and indeed biology as a whole. The three crucial problems were: gene expression, variation and function. He illustrated the rapid increase in the number of publications on the subject since mid 1995. Studies in yeast were likely to show advances in the field, because of the smaller genome. He emphasized that microarray studies should be correlated with clinical studies. Rigorous statistical analysis and formal mathematical modeling would be necessary.

Using microarray data and artificial neural networks, he presented information to distinguish estrogen receptor positive from negative breast cancer tumors. He emphasized that inaccuracy may result from measurement variations and from biological variables. The latter could be more important. Array data is colour characterized, but the human eye has limited ability to distinguish colour variations. Genes are regulated across multiple conditions, and there could be a hierarchical clustering of hormone regulated genes.

Speaking on 'GnRH frequency dependence studied by high throughput genomics,' Stuart C Sealfon began that molecular biology is now about the same stage as astronomy was in Babylonian times. Progress can be expected from an integration of genomic and computational processes. He stated that intrinsic genes can be used as a monitor of cell signaling, although many points may be necessary for time trajectory information. While microarrays have advantages such as ability to assay large number of genes at low cost, its limitations must be considered: poor data quality, incorrect gene assignment and low sample throughput. Essentially microarrays have potentially low sensitivity. To overcome these, real time PCR must confirm and validate array data. Using currently available instrumentation, robotic arms can perform a far larger number of assays. Kinetic PCR, which may be considered the gold standard, has a sensitivity of 85-90%.

In the GnRH coupled gene network, 67 genes were induced by six hours. Gene network and signal transduction formed a continuum, with feedback from genome to signaling. The number of induced genes decreased over time. To add to the complexity, behavior of single cell may be different from that of a group of cells. Feed forward loop is tightly regulated. The modular servomechanism may be summarized as being feed forward serial, recurrent serial, feed forward parallel and recurrent parallel. He concluded that signaling network and gene network in the GnRH system form a continuous cell replication in a simple manner. However, as this occurs at different sites and at different times, complexity results, analogous to fractals and the chaos phenomenon.

Taking a practical application of microarrays, Katharyn B Horwitz elaborated on 'cDNA arrays and hormone responsiveness in breast cancer.' Microarrays were employed to study sex hormones and the biology of breast cancer. In women, estrogens promote growth of breast carcinoma cells, but do not initiate its development. Hormone antagonism offers a valuable method of treatment. Oophorectomy, and drugs such as tamoxifen, naloxifine and fulvestrat are used. Following the recent evidence of adverse outcome in the trial with estrogen-progesterone, attention was focused on two subgroups of

progesterone receptor: the larger B form and the smaller A form. In the presence of progesterone, PR-B regulates more genes than PR-A. Practical fallout would be screening for progesterone receptor subtypes. Could PR-A specific progesterone agonists be safely used in HRT? A recent study showed that estrogen dependent tumor growth was preferentially inhibited by PR-A. Genes regulated by PR-A and B were identified by microarray technology. However they cannot explain tissue specific receptor mechanism of action. Utilizing newer microarray techniques, more clinically relevant information can be obtained on hormone receptor status in breast cancer.

Turning attention to the placenta, Stuart Handweiger spoke on the 'Categorical reprogramming of genes during differentiation of the human placenta and uterus.' Two principal reasons for studying placental development are, abnormal placenta results in early fetal loss, and the role of placental abnormalities in pregnancy induced hypertension.

The placenta is a complex organ which functions in structural integrity, gas transfer, excretion, hormone production, hematopoiesis and immune function. Most of these are done at the level of placental villi, which, like the lung alveolus has a large surface area. Villous trophoblast differentiates from cytotrophoblast to syncitiotrophoblast through both morphological and biochemical differentiation. The critical questions in this process are, what genes induce/repress the differentiation, and the pattern of gene expression. Using microarray (Incyte Human Gem/ microarray) with 697 genes, primary data were examined using Incye Gentools and GeneSpring software. One hundred forty one genes were induced two fold or more, and 256 genes repressed two fold or more. Given the limitations of genes represented on the chip, one must remember that an important gene such as hCG was not represented.

During trophoblast differentiation, multiple patterns and functional classes of genes were identified in structural dynamics, cell cycle, apoptosis, intercellular communication and intracellular metabolism. The earliest genes that were dynamically regulated were involved in cell adhesion and in fusion. Genes regulated later were principally involved in differentiation function such as hormone production, intermediary metabolism and substrate transport.

On the second day of the symposium, Klaus H Kaestner began the proceedings with an account of 'Functional genomics of the endocrine pancreas: the pancreas clone set and PancChip, new resources for diabetes research.' PancChip was the outcome of the Functional Genomics Consortium, which aims to assemble pilot microarrays to construct and sequence endocrine pancreas enriched cDNA libraries and to establish Endocrine Pancreas Consortium database, while giving support in bioinformatics. An experimental and computational hybrid was synthesized from commercially available chips, with cluster ESTs placed contiguously. In the Pancreas Chip 2.0, approximately 3,400 clones were represented, with 3,139 unique clones. There were 231 clones representing pathways known to be important in the endocrine pancreas, and 30 housekeeping genes. Using pancreas RNA from fetal tissue, hybridization to PancChip 2.0 identified sets of 'early response' and 'late response' genes.

Critical to maintaining data fidelity, libraries were screened for quality of data. Poor data set libraries were discarded. To data the mouse Consortium library, from 38,660 mouse databank, there were 9,464 transcriptomes (ie unique mRNAs), with 1,821 novel sequences identified for the first time. In the human consortium library there were 13,910 transcriptomes, with 2,529 novel sequences. In collaborative exercises across scientific disciplines, Kaestnor cautioned, it was essential to discuss with the biostatistician even before the first RNA was analyzed.

The EPConDB contains pancreas clone set, microarray results and genes expressed in the pancreas. Information can be obtained by Boolean queries, which can be archived, with downloadable microarray data.

Referring to practical applications of DNA sequencing, Andrew Hattersley spoke on 'The DNA sequencer in the diabetic clinic.' He began with an arresting question: who would you rather consult at the diabetic clinic? On one side was a clinician and on the other a complex DNA sequencer. The diagnosis of diabetes that depended on clinical and biochemical criteria alone is no longer sufficient. Recently a number of monogenic diabetic genes were identified. Among them, mutations of glucokinase enzyme and transcription factor lead to maturity onset diabetes of young. The mutations are not only useful to understand the etiology, but have prognostic value: in glucokinase mutations, diabetic complications are uncommon unlike transcription factor mutations.

These can be identified using DNA sequencer, which are generally available in the United Kingdom.

However ethical questions come up such as who decides and whether testing must be done, besides technological limitations. Besides the cost of each test is substantial (UK Pounds 300/- a test). The availability of DNA sequences must be used with appropriate clinical skills in application and interpretation of test results. In the final slides, he answered the teaser projected at the beginning, that clinician and DNA sequencer are both necessary.

In the concluding session, Mike Moran began with 'Applications of mass spectrometry towards clinical proteomics.' He focused on three aspects: protein interaction for pathway mapping, new technology platforms and protein expression as related to phosphorylation. Basically proteins function through domains. Protein-protein interactions can be studied by preliminarily assigning a function, followed by pathway identification.

In mining target classes, specifically phosphorylative networks, use of sensitive detectors was crucial: hybrid quadrapole ion trap Fourier transform ICR mass spectrometer had low chemical noise, better mass accuracy and dynamic range. Phosphoprofile was demonstrated as a measure of cellular status. In future bioinformatic dependent analysis can be done using clinical tissue samples. Using differential analysis, protein expression in compartments can be done with identification of disease specific cell surface receptors.

Carrying on the theme of mass spectrometry, Andrew Emili addressed 'The proteomics of endocrinology: mass spectrometry in overdrive.' His focus was on identifying endocrine networks, and protein function using mass spectrometry. At the outset he said proteomics and conventional protein biochemistry differed in scale and scope. Proteomics deals with protein activity, circuits, expression, turnover, localization and modification. Presently, abundant genomic data exist, with little functional annotation about protein function downstream. Such annotations can be obtained by the concept of 'guilt by association.' Proteins occurring together are likely to have similar functions. Tandem mass spectrometry can rapidly identify proteins. Multiple data points are compared with reference databases. Technical problems can result from limitations in the quality of reference databases. A next logical step would be expression proteomics, which answers questions such as how does a protein relate to cell physiology, development and disease?

Ultimately database mining is a compromise between accuracy (statistical) and sensitivity (technical). He briefly described PRISM that simplifies data sets to make them accessible. The GO is a dynamic controlled vocabulary of protein function, whose power increases as the database is updated with new information. The GOClust covers major biological processes in liver and lung. From available evidence, it appears to have biological relevance. He concluded that progress depends on close collaboration with groups of scientists competent in different technologies.

The last presentation was made by Isaac S Kohane on 'Bioinformatics for an integrated endocrinology.' He stated that major problems awaiting solutions for an integrative endocrinology were dynamic aspects (ie importance of time), integration of genotypic and phenotypic variables, and ultimately an incorporation of functional genomics into treatment.

Clustering is a statistical model selection procedure. Temporal order gives insights into concerted behavior of gene expression. He emphasized the necessity of sound statistical basis for models, rather than 'looks-right' tests. Frameworks, such as Bayesian network-framework for probabilistic relationship, may do an integration of disparate genotype-phenotype data sets.

Functional genomics in prognosis may in turn have its own challenges; eg, which tissue must be used for expression? In diabetes, should the liver, fat, brain or all be used, and why? He concluded that a new integrative course was instituted at Harvard University and Massachusetts Institute of Technology for students who graduated.

In addition to lectures, six free papers were presented with data on identification of new candidate genes in sporadic macronodular adrenal hyperplasia, early glucocorticoid genes in apoptopic eosinophils, identification of FXR target genes, gene expression in human parathyroid tumors, in metastatic insulinomas and gene expression pattern identification using computational bioinformatics.

In summary, the symposium offered a platform for looking back at the human genome sequencing project, and critically, looking ahead. The contours of the grand scope of biology were discerned, areas of ignorance acknowledged, and a fascinating look into the present and future of integrative biology was on display.