MSRI Mathematical Sciences Research Institute

Mathematical Systems Biology of Cancer

Organized By: Dick Karp, Bahram Parvin, Terry Speed, Paul Spellman, Carolyn Talcott, Wing Wong
May 3, 2006 to May 5, 2006

Lecture Schedule

Wednesday, May 3

8:45 – 9:00 *Welcome by MSRI*

Three statistics tutorials

9:00 – 9:40 **Jane Fridlyand** (Epidemiology & Biostatistics, UCSF): "Analysis of gene expression and CGH microarray data"

Abstract: This tutorial will consist of the two parts. In the first part we will describe Microarray-based Comparative Genomic Hybridization (Array CGH) which is a technique that measures DNA copy number changes, and localizes them on the genome. Such copy number aberrations are common in cancer and in many developmental abnormalities. In the second part we will introduce Affymetrix transcriptional profiling technology. After outlining both approaches to genomic profiling, we will discuss statistical methods currently used for their data normalization and analysis. Panel of the 60 breast cancer cell lines will be used for illustrative purposes.

9:40 – 10:10 **Keith Baggerly** (Biostatistics, M D Anderson Cancer Center): "Analysis of reverse-phase protein lysate arrays"

Abstract: In recent years, high-throughput biological assays have been primarily focused on RNA (microarrays). In this talk, we will describe how people are attempting to extend the high-throughput paradigm to the exploration of the human proteome. Much of this talk is focused on trying to give a clear explanation of how the various types of proteomic assays we have fit together, and how measuring proteins is different from measuring DNA or RNA. To that end, I'm walking through the basics of ELISAs, Western blots, and arrays, simply so that the combination in a reverse-phase lysate array can be seen as an evolution of several processes. Having placed the assays in context, we will look at some data generated at MD Anderson, and try to address a basic analysis question: how do people quantify lysate array data? The hybridization kinetics on a reverse phase array are more similar across an array than is the case with standard microarrays, so that borrowing strength across samples within an array is more feasible. Borrowing across arrays, by contrast, is less so.

10:10 –10:30 **Frances Tong** (Statistics, UCB): "Normalization of Western blots"

Abstract: One common method to elucidate the presence and levels of proteins in cancer cell lines is western blotting. Also known as immunoblots, they detect proteins in a given sample of tissue with separation by gel electrophoresis and then probed by antibodies specific to the protein. This talk will discuss the normalization of numerous sets of western blots on breast cancer cell lines from the Gray lab. Each protein is detected across 38 breast cancer cell lines in a set of 4 blots. To enable comparison across blots, a pair of cell lines are common in each blot. The samples are adjusted on a multiplicative scale accordingly to the relationships between the common samples.

10:30 – 11:00 *Tea Break*

11:00 – 12:00 **Joe Gray** (Life Sciences, LBNL): "Predicting response to pathway targeted cancer therapies; an ICBP systems opportunity"

12:00 –1:30 *Lunch*

1:30 – 2:30 **Claire Tomlin** (Aeronautics & Astronautics, Stanford University): "Hybrid system models for protein regulatory networks - Models and analysis methods. Application to Notch Delta signaling"

Abstract: In this talk, methods that we have designed to analyze and help to identify certain protein regulatory networks will be presented. Hybrid automata represent one suitable modeling framework, as the protein concentration dynamics inside each cell are modeled using linear differential equations; inputs activate or deactivate these continuous dynamics through discrete switches, which themselves are controlled by protein concentrations reaching given thresholds. We present an iterative refinement algorithm for computing discrete abstractions of a class of symbolic hybrid automata, and we apply this algorithm to a model of multiple cell Delta-Notch protein signaling. The results are analyzed to show that novel, non-intuitive, and biologically interesting properties can be deduced from the computation, thus demonstrating that mathematical modeling which extrapolates from existing information and underlying principles can be successful in increasing understanding of some biological systems. Joint work with Ronojoy

Ghosh and Jeff Axelrod.

2:30 – 3:30 **Rich Neve** (Life Sciences, LBNL): "The ICBP data set; an EPHA2 example"
Abstract: Many hurdles need to be negotiated when interpreting biologic data in terms of mathematical models. Not least of these is the requirement for biologists and mathematicians to understand each others semantics in order to build the correct models. I will present a broad overview of how biologists use data to build pathway models and explain the complexities that exist when considering biologic processes in terms of models.

3:30 – 4:00 *Tea Break*

4:00 – 5:00 **Claire Tomlin** (Aeronautics & Astronautics, Stanford University): "Using the adjoint method for parameter identification of large scale protein regulatory networks. Application to planar cell polarity in Drosophila"

Thursday, May 4

8:30 – 9:30 **Paul Spellman** (LifeSciences, LBNL): "Pathway modelling"

9:15 – 10:00 **Paraic Kenny** (Life Sciences, LBNL): "3D models of normal and malignant breast epithelial cells"
Abstract: Culturing cells ex vivo has led to a substantial increase in our understanding of cellular processes over the past half century. However, cells cultured on flat, two-dimensional plastic substrata rapidly lose many aspects of the differentiated phenotype and behave, in many respects, differently to cells in vivo. The Bissell laboratory has pioneered the use of three-dimensional cultures of cells in gels consisting of the basement membrane proteins to which these cells adhere in vivo. Using these techniques we have succeeded in maintaining the differentiated phenotype in mouse and human mammary epithelial cells and recapitulate much of the complex architecture of the mammary gland. Furthermore, we show that these assays effectively distinguish malignant from non-malignant cells. I will discuss the application of 3D culture techniques in cancer biology, emphasizing the important contributions these can make to understanding treatment response, identifying key molecular targets for drug treatment and prediction of patient outcome.

10:00 – 10:45 **Michael Korn** (Cancer Center USCF): “CAR mediated signalling”
Abstract: Novel therapeutic agents specifically targeting pathways involved in cancer pathogenesis have been developed and are promising components of rationally designed combination treatments for cancer. Theoretically, such therapies could be designed to target specifically the molecular composition of individual tumors. We are pursuing a systems-based analysis approach to move towards this goal. Large-scale measurements of molecular and cellular responses to targeted interventions into cancer-relevant signal transduction pathways are being performed and used to construct mathematical pathway models. Our aim is to use such models for predictions of molecular responses to targeted therapies and to reveal nodes within signal transduction networks that render cancer cells particular susceptible to therapeutic interventions.

10:45 – 11:15 *Tea Break*

11:15 – 12:15 **Carolyn Talcott (**Computer Science, SRI International): "Pathway Logic Models"
Abstract: Pathway Logic is an approach to modeling cellular processes based on symbolic logic. It allows one to model aspects of the structure and state of interacting components, to represent individual process steps (reactions) and to study possible ways a system could evolve using techniques based on logical inference. Reactions can be modeled at many levels of detail ranging from micro steps representing events such as phosphorlyation at specific sites or binding of protein domains to macro steps such as the results of signaling or metabolic modules. Given a network of reactions and a specification of cellular components one can use logical inference to query the network about possible reaction pathways and outcomes. In this talk we will explain how signal transduction reactions are represented in Pathway Logic, and illustrate some of the questions that one can ask a model using the Pathway Logic Assistant. Then we will describe several ways in which a Pathway Logic model can be used to analyze gene expression data.

12:15 – 1:45 *Lunch*

1:45 – 2:30 **Mary Helen Barcellos-Hoff** (Life Sciences, LBNL): "Integrative Radiation Biology"
Abstract: In the context of multicellular organisms, an orchestrated response to DNA damage by ionizing radiation is important for rapid restoration of homeostasis and long-term prevention of cancer. The challenge in predicting radiation health effects in humans is to understand how cellular responses occurring in a multicellular context are integrated to produce an organismal response. Experimental studies show that radiation exposure elicits responses that can produce effects in non-irradiated bystander cells or can lead to a high frequency of genomic instability in the progeny of irradiated cells. This has motivated a substantial effort to both describe and quantify

these non-targeted responses. One may argue that, more importantly, those data have heightened awareness that many types of cell interactions contribute to long term radiation effects, and that multicellular responses are poorly integrated into the current paradigms of radiation effects and their consequences in terms of human health. A model of radiation response based on the systems biology principles of network interconnectivity and spatial organization is discussed that would reconcile the apparent contradiction of these cellular responses within the higher order structure of tissues and organisms.

2:30 – 3:30 **Mike West (**Statistics & Decision Sciences, Duke University): "Data, Models and Computation in the Duke NCI Integrative Cancer Biology Program"

Abstract: I will discuss some of the activities in the projects linked to the Duke Integrated Cancer Biology Program - one of the NCI - ICBP centers. The Duke ICBP projects concern data, methods, models and analyses that aim to improve our understanding of the complex Rb/E2F network. This network of interconnected oncogenic signaling pathways (including the Ras, Myc, Rb-E2F and p53 response pathways) is fundamental to the control of cell cycle, links the activity of cellular proliferation processes with the determination of cell fate, and is subject to myriad aspects of deregulation that relate to the development of human cancer.

The ICBP projects include the development of multiple forms of molecular data (gene expression, protein interaction, DNA binding, metabolic) as well as the development and application of statistical analysis methods and other modelling components including emerging work on single cell dynamics connected to the cell cycle. Some of the statistical focus to date has focused heavily on gene expression studies; these play into the basic studies that aim to improve characterization of complex patterns of pathway deregulation at a fundamental level, as well as the ensuing applied studies that aim to evaluate the resulting information in clinical outcomes studies. Some of the statistical work involves sparse regression and factor models that couple modelling approaches relevant to both goals, and I'll discuss some of the underlying ideas, modelling details, computation and examples of these models (which are of course of general, and are being applied in a number of other areas too).

3:30 – 4:00 *Tea Break*

4:00 – 5:00 **Bud Mishra** (Computer Science, NYU): “Inheritance of Loss: Computational Systems Biology for Cancer”

Abstract: Rapid and accurate solutions to many biomedical problems are beginning to rely on systems and computational approaches. Few notable examples are: genomic assays for cancer, genetic analysis of cancer genomes for marker detection, models of cancer progression, etc. While these examples focus on cancer and currently build upon microarray technology, the algorithmic approaches must aim to be scalable, agnostic to the technologies, and applicable to a wide variety of problems. This talk surveys many promises, challenges and obstacles faced by the emerging field of systems biology as they tackle these biomedical problems. In particular, we will emphasize three highly-intertwined aspects of this problem:

** Measuring: Array and Single-Molecule Measurements for More Informative Data*
** Mining: Combining Gene Expression and Genomic Patient Data for Discovery*
** Modeling: Systems Biology algorithms for reasoning and redescription of time-course data.*

Various novel applications of mathematical ideas appear: 0-1-Laws in experiment design, Hidden Kripke Models and temporal logic redescriptins, Nonlinear Kalman-Bucy filtering, Efficient Maximum A Posteriori Estimators, etc.

Friday, May 5

8:30 – 9:15 **Damir Sudar** (Life Sciences, LBNL): "Computational bioimaging and informatics, I"

9:15 – 10:00 **Bahram Parvin** (Life Sciences, LBNL): "Computational bioimaging and informatics, II"

Abstract: Organisms express their genome in a cell-specific manner, resulting in a variety of cellular phenotypes or phenomes. Mapping cell phenomes under different experimental conditions is necessary for understanding and modeling the responses of organisms. From the experimental perspective, such mapping requires an effective model system and assays that facilitate visualization of each specific endpoint. From the computational perspective, such mapping requires detailed annotations that are tightly coupled with computed quantitative representation for subsequent modeling and comparative analysis. In this talk, instrumentation, computational and informatics techniques are presented to meet the challenges of high content analysis for data collected through each mode of optical microscopy and each specific assay. In order to systematically collect image data for high-throughput analyses, required for systems biology, imaging systems will be described that permit or enforce highly structured image acquisition.

Considering the importance of understanding the imaging (and specimen preparation) procedures in order to properly analyze the data, significant emphasis will be put on the details of image formation. Computational methods capture a spatial configuration of pertinent events, such as localization or binding efficiency, through an undirected attributed graph representation, while the informatics framework provides the foundation for detailed annotation, experimental design, robotics interface, and organization of data for subsequent comparative and multiscale analysis. One critical component is to leverage and couple annotation and corresponding data with emerging new standards in imaging and bioinformatics. Another component is effective visualization of data and computed representation.

10:00 – 10:45 **Larry Lok** (Molecular Sciences Institute, Berkeley): "Optimal binning strategies"
Abstract: At the Molecular Sciences Institute and all over, researchers apply discrete Bayesian networks to model intracellular processes, including gene regulation and signal transduction. I will note a few ways of binning continuous data, such as come from flow cytometry, in order to use discrete statistical tools. I will review some information-theoretic background on two binning strategies and present some computational comparisons.

10:45 – 11:15 *Tea Break*

11:15 – 12:15 **Trey Ideker** (Bioengineering UCSD): "Protein network comparative genomics"
Abstract: With the appearance of large networks of protein-protein and protein-DNA interactions as a new type of biological measurement, methods are needed for constructing cellular pathway models using interaction data as the central framework. The key idea is that, by comparing the molecular interaction network with other biological data sets, it will be possible to organize the network into modules representing the repertoire of distinct functional processes in the cell. Three distinct types of network comparisons will be discussed, including those to identify:
(1) Protein interaction networks that are conserved across species
(2) Networks in control of gene expression changes
(3) Networks correlating with systematic phenotypes and synthetic lethals
Using these computational modeling and query tools, we are constructing network models to explain the physiological response of yeast to DNA damaging agents.
Relevant articles and links:
*Yeang, C.H., Mak, H.C., McCuine, S., Workman, C., Jaakkola, T., and Ideker, T. Validation and refinement of gene regulatory pathways on a network of physical interactions. Genome Biology **6(7):** R62 (2005).*
*Kelley, R. and Ideker, T. Systematic interpretation of genetic interactions using protein networks. Nature Biotechnology **23(5):**561-566 (2005).*
*Sharan, R., Suthram, S., Kelley, R. M., Kuhn, T., McCuine, S., Uetz, P., Sittler, T., Karp, R. M., and Ideker, T. Conserved patterns of protein interaction in multiple species. Proc Natl Acad Sci U S A. **8:102(6):** 1974-79 (2005).*
*Suthram, S., Sittler, T., and Ideker, T. The Plasmodium network diverges from those of other species. Nature **437:** (November 3, 2005).*
http://www.pathblast.org
http://www.cytoscape.org
***Acknowledgements:** We gratefully acknowledge funding through NIH/NIGMS grant GM070743-01; NSF grant CCF-0425926; Unilever, PLC, and the Packard Foundation.*

12:15 – 1:30 *Lunch*

1:30 – 2:30 **Jasmine Zhou** (Molecular and Computational Biology USC): "Identifying cancer genetic network signature from integrative microarray analysis"
Abstract: Microarray gene expression profiling has been widely applied to cancer research. The commonly used analysis approach is to identify genes differentially expressed in cancer versus normal tissues or those in different cancer subtypes. However, it is known that phenotypes are determined not only by genes, but also by the underlying structure of genetic networks. Often, it is the interaction of many genes that causes phenotypic differences. In this work, we develop graph-based methods to integrate multiple microarray data sets for cancer-related network module discovery. By transforming each microarray dataset into a co-expression network, we perform comparative network analysis to extract subnetworks which occur frequently in cancer datasets but not in other datasets. In this way, we identify network modules that characterize commonalities and variances of different types of cancer. Finally, we focus on dense co-expression networks which represent cancer-specific co-expression clusters, and assign putative transcriptional regulator to them.

2:30 – 3:30 **Ben Raphael** (Bioengineering UCSD): "Analysis of large-scale alterations in tumor genomes"

Abstract: Cancer is a disease driven by mutations in the genome that alter the structure, function, and regulation of genes. These mutations range from single letter changes in the DNA sequence to more drastic rearrangements, gains, or losses of large pieces of DNA. In some types of cancer these large-scale alterations are directly implicated in the pathogenesis of cancer and provide targets for cancer diagnostics and therapeutics.

I will describe computational methods for reconstructing tumor genome architectures and analyzing rearrangements in tumor genomes at high resolution using a technique called End Sequence Profiling (ESP). These methods produce a parsimonious sequence of rearrangements that transform the normal human genome into a tumor genome. Furthermore, computational analysis of ESP data suggests mechanisms that produce complicated patterns of overlapping rearrangement and duplication events that are observed in some tumor genomes. Another experimental technique called array comparative genomic hybridization (aCGH) has become indispensable in the identification of duplicated and deleted segments of DNA in tumor genomes. ESP provides an effective complement to aCGH, and I will discuss how to combine data from both types of experiments to obtain a comprehensive view of tumor genome architecture.

3:30 – 4:00 *Tea Break*

4:00 – 5:00 **Steve Ethier** (Karmanos Cancer Institute) & **Greg Dewey** (Keck Graduate Institute): "Network Analysis of Gene Expression Kinetics in Human Breast Cancer Cells"

Abstract: Experiments were designed to examine the gene expression networks regulated by the epidermal growth factor receptor (EGFR) in the normal human mammary epithelial cell line MCF-10A, and in the human breast cancer cell line, SUM-149. These cell lines were chosen due to their absolute dependence on EGFR signaling for growth in vitro. To generate gene expression networks from the microarray data, we used a simple dynamic model to analyze the time response of the system following the inhibition of EGFR activity using small molecule tyrosine kinase inhibitors specific for EGFR. The model is based upon a linear finite difference equation in which the estimation of a transition matrix led us to establish the topology of the underlying network. Local measures of connectivity like weighted connectivity and clustering coefficient allowed us to identify characteristic major hubs from the breast cancer cell line. By defining a particular threshold, we were able to convert the original weighted matrix to an adjacency matrix in which global measures of connectivity can be applied. One of these measures, the mutuality or reciprocity, suggested a cooperation decrease between major hubs in the cancer cell line as compared to the normal cells. In addition, by comparing network topology between normal mammary epithelial cells and breast cancer cells that both rely on EGFR signaling for growth and survival, we were able to identify cancer-cell specific hub genes within the EGFR-regulated networks. Whereas some of the hub genes, such as DUSP6 and AREG, would have been predicted based on prior knowledge of EGFR signaling pathways in normal and neoplastic cells, many of the cancer cell specific hubs would not have been predicted to be regulated by EGFR. Among these hub genes are IL-1A, IL-1B and genes related to cytokine signaling and NF-kappaB activation. We confirmed that expression of these cytokines was regulated by EGFR in the breast cancer cells and not in the normal cells, and further demonstrated that blockade of these pathways leads to growth arrest and loss of viability in the cancer cells but not in normal cells. Thus, kinetic analysis of the regulation of gene expression networks in cancer cells and normal cells has the potential to identify novel features of cell signaling and gene expression in cancer cells, which in turn, can predict novel targets for therapeutic intervention.