Summary of Articles

Daniel Hyduke, Lars Rohlin, Katy Kao and James Liao. A Software Package for cDNA Microarray Data Normalization and Assessing Confidence Intervals.

DNA microarray data are affected by variations from a number of sources. Before these data can be used to infer biological information, the extent of these variations must be assessed. Here we describe an open source software package, lcDNA, that provides tools for filtering, normalizing, and assessing the statistical significance of cDNA microarray data. The program employs a hierarchical Bayesian model and Markov Chain Monte Carlo simulation to estimate gene-specific confidence intervals for each gene in a cDNA microarray data set. This program is designed to perform these primary analytical operations on data from two-channel spotted, or *in situ* synthesized, DNA microarrays.

Rajanikanth Vadigepalli, Praveen Chakravarthula, Daniel Zak, James Schwaber and Gregory Gonye. PAINT: A Promoter Analysis and Interaction Network Generation Tool for Gene Regulatory Network Identification.

We have developed a bioinformatics tool named PAINT that automates the promoter analysis of a given set of genes for the presence of transcription factor binding sites. Based on coincidence of regulatory sites, this tool produces an interaction matrix that represents a candidate transcriptional regulatory network. This tool currently consists of: (1) a database of promoter sequences of known or predicted genes in the Ensembl annotated mouse genome database, (2) various modules that can retrieve and process the promoter sequences for binding sites of known transcription factors, and (3) modules for visualization and analysis of the resulting set of candidate network connections. This information provides a substantially pruned list of genes and transcription factors that can be examined in detail in further experimental studies on gene regulation. Also, the candidate network can be incorporated into network identification methods in the form of constraints on feasible structures in order to render the algorithms tractable for large-scale systems. The tool can also produce output in various formats suitable for use in external visualization and analysis software. In this manuscript, PAINT is demonstrated in two case studies involving analysis of differentially regulated genes chosen from two microarray data sets. The first set is from a neuroblastoma N1E-115 cell differentiation experiment, and the second set is from neuroblastoma N1E-115 cells at different time intervals following exposure to neuropeptide angiotensin II. PAINT is available for use as an agent in BioSPICE simulation and analysis framework (http://www.biospice.org), and can also be accessed via a WWW interface at http://www.dbi.tju.edu/dbi/tools/paint/.

Bud Mishra, Raoul-Sam Daruwala, Yi Zhou, Nadia Ugel, Alberto Policriti, Marco Antoniotti, Salvatore Paxia, Marc Rejali, Archisman Rudra, Vera Cherepinsky, Naomi Silver, William Casey, Carla Piazza, Marta Simeoni, Paolo Barbano, Marina Spivak, Jiawu Feng, Ofer Gill, Mysore Venkatesh, Fang Cheng, Bing Sun, Iuliana Ioniata, Thomas Anantharaman, E. Jane Albert Hubbard, Amir Pneuli, David Harel, Vijay Chandru, Ramesh Hariharan, Michael Wigler, Frank Park, Suih-Chieh Lin, Yuri Lazebnik, Franz Winkler, Charles Cantor, Alessandra Carbone and Mikhael Gromov. A Sense of Life: Computational & Experimental Investigations with Models of Biochemical & Evolutionary Processes.

We collaborate in a research program aimed at creating a rigorous framework, experimental infrastructure, and computational environment for understanding, experimenting with, manipulating, and modifying a diverse set of fundamental biological processes at multiple scales and spatio-temporal modes. The novelty of our research is based on an approach that (i)

requires coevolution of experimental science and theoretical techniques and (ii) exploits a certain universality in biology guided by a parsimonious model of evolutionary mechanisms operating at the genomic level and manifesting at the proteomic, transcriptomic, phylogenic, and other higher levels. Our current program in \systems biology" endeavors to marry large-scale biological experiments with the tools to ponder and reason about large, complex, and subtle natural systems. To achieve this ambitious goal, ideas and concepts are combined from many different fields: biological experimentation, applied mathematical modeling, computational reasoning schemes, large-scale numerical and symbolic simulations, etc. From a biological viewpoint, the basic issues are many: (i) understanding common and shared structural motifs among biological processes; (ii) modeling biological noise due to interactions among a small number of key molecules or loss of synchrony; (iii) explaining the robustness of these systems in spite of such noise; and (iv) cataloging multistatic behavior and adaptation exhibited by many biological processes.

P. Ortoleva, E. Berry, Y. Brun, J. Fan, M. Fontus, K. Hubbard, K. Jaqaman, L. Jarymowycz, A. Navid, A. Sayyed-Ahmad, Z. Shreif, F. Stanley, K. Tuncay, E. Weitzke and L.-C. Wu. The Karyote Physico-Chemical Genomic, Proteomic, Metabolic Cell Modeling System.
Modeling approaches to the dynamics of a living cell are presented that are strongly based on its underlying physical and chemical processes and its hierarchical spatio-temporal organization. Through the inclusion of a broad spectrum of processes and a rigorous analysis of the multiple scale nature of cellular dynamics we are attempting to advance cell modeling and its applications. The presentation focuses on our cell modeling system that integrates data archiving and quantitative physico-chemical modeling and information theory to provide a seamless approach to the modeling/data analysis endeavor. Thereby the rapidly growing mess of genomic, proteomic, metabolic and cell physiological data can be automatically used to develop and calibrate a predictive cell model.

The discussion focuses on the *Karyote*® cell modeling system and an introduction to the *CellX*® and *VirusX*® models. The *Karyote* software system integrates three elements:

- a model-building and data archiving module that allows one to define a cell type to be modeled through its reaction network, structure, and transport processes as well as to choose the surrounding medium and other parameters of the phenomenon to be modeled;
- a genomic, proteomic, metabolic cell simulator that solves the equations of metabolic reaction, transcription/translation polymerization and the exchange of molecules between parts of the cell and with the surrounding medium; and
- an information theory module (ITM) that automates model calibration and development and integrates a variety of data types with the cell dynamics computations.

In *Karyote*, reactions may be fast (equilibrated) or slow (finite rate) and the special effects of enzymes and other minority species yielding steady-state cycles of arbitrary complexities are accounted for. These features of the dynamics are handled via rigorous multiple scale analysis. A user interface allows for an automated generation and solution of the equations of multiple timescale, compartmented dynamics.

Karyote is based on a fixed intracellular structure. However, cell response to changes in the host medium, damage, development or transformation to abnormality can involve dramatic changes in intracellular structure. As this changes the nature of the cellular dynamics, a new model, *CellX*, is being developed based on the spatial distribution of concentration and other variables. This allows *CellX* to capture the self-organizing character of cellular behavior. The self-assembly of organelles, viruses and other subcellular bodies is being addressed in a second new model, *VirusX*, that integrates molecular mechanics and continuum theory. *VirusX* is designed to study the influence of a host medium on viral self-assembly, structural stability, infection of a single cell and transmission of disease.

Nicholas Allen, Laurence Calzone, Katherine Chen, Andrea Ciliberto, Naren Ramakrishnan, Clifford Shaffer, Jill Sible, John Tyson, Marc Vass, Layne Watson and Jason Zwolak. Modeling Regularoty Networks at Virginia Tech.

The life of a cell is governed by the physicochemical properties of a complex network of interacting macromolecules (primarily genes and proteins). Hence, a full scientific understanding of and rational engineering approach to cell physiology require accurate mathematical models of the spatial and temporal dynamics of these macromolecular assemblies, especially the networks involved in integrating signals and regulating cellular responses. The Virginia Tech Consortium is involved in three specific goals of DARPA's computational biology program (Bio-COMP): to create effective software tools for modeling gene-protein-metabolite networks, to employ these tools in creating a new generation of realistic models, and to test and refine these models by well-conceived experimental studies. The special emphasis of this group is to understand the mechanisms of cell cycle control in eukaryotes (yeast cells and frog eggs). The software tools developed at Virginia Tech are designed to meet general requirements of modeling regulatory networks and are collected in a problem-solving environment called JigCell.

Daniel Segre, Jeremy Zucker, Jeremy Katz, Xiaoxia Lin, Patrick D'haeseleer, Wayne Rindone, Peter Karchenko, Dat Nguyen, Matthew Wright and George Church. From annotated genomes to metabolic flux models and kinetic parameter fitting.

Significant advances in system-level modeling of cellular behavior can be achieved based on constraints derived from genomic information and on optimality hypotheses. For steady state models of metabolic networks, mass conservation and reaction stoichiometry impose linear constraints on metabolic fluxes. Different objectives, such as maximization of growth rate or minimization of flux distance from a reference state, can be tested in different organisms and conditions. In particular, we have suggested that the metabolic properties of mutant bacterial strains are best described by an algorithm that performs a minimization of metabolic adjustment (MOMA) upon gene deletion. The increasing availability of many annotated genomes paves the way for a systematic application of these flux balance methods to a large variety of organisms. However, such a high throughput goal crucially depends on our capacity to build metabolic flux models in a fully automated fashion. Here we describe a pipeline for generating models from annotated genomes, and discuss the current obstacles to full automation. In addition, we propose a framework for the integration of flux modeling results and high throughput proteomic data, which can potentially help in the inference of whole-cell kinetic parameters.

Chris Cox, Gregory Peterson, Michael Allen, Joseph Lancaster, James McCollum, Derek Austin, Ling Yan, Gary Sayler and Michael Simpson. Analysis of Noise in Quorum Sensing.

Noise may play a pivotal role in gene circuit functionality, as demonstrated for the genetic switch in the bacterial phage λ. Like the λ switch, bacterial quorum sensing (QS) systems operate within a population and contain a bistable switching element, making it likely that noise plays a functional role in QS circuit operation. Therefore, a detailed analysis of the noise behavior of QS systems is needed. We have developed a set of tools generally applicable to the analysis of gene circuits, with an emphasis on investigations in the frequency domain (FD), that we apply here to the QS system in the marine bacterium *Vibrio fischeri*. We demonstrate that a tight coupling between exact stochastic simulation and FD analysis provides insights into the structure/function relationships in the QS circuit. Furthermore, we argue that a noise analysis is incomplete without consideration of the power spectral densities (PSDs) of the important molecular output signals. As an example we consider reversible reactions in the QS circuit, and show through analysis and exact stochastic simulation that these circuits make significant and dynamic modifications to the noise spectra. In particular, we demonstrate a ‘whitening’ effect, which occurs as the noise is processed through these reversible reactions.

Thomas Garvey, Patrick Lincoln, Charles John Pedersen, David Martin and Mark Johnson. BioSPICE Access to biological computation for analysis and design.

The goal of the BioSPICE program is to create a framework that provides biologists access to the most current computational tools for analysis and design of experiments. At the program midpoint, BioSPICE comprises software contributions from approximately 20 participating laboratories integrated under the BioSPICE Dashboard, and a methodology for continued software integration. These contributed software modules are the BioSPICE Dashboard, a graphical environment that combines Open Agent Architecture and NetBeans software technologies in a coherent, biologist-friendly user interface. The current Dashboard permits data sources, models, simulation engines, and output displays provided by different investigators and running on different machines to work together across a distributed, heterogeneous network. Among several other features, the Dashboard enables users to create graphical workflows by configuring and connecting available BioSPICE components. Anticipated future enhancements to BioSPICE include a notebook capability that will permit researchers to browse and compile data to support model building, a biological model repository, and tools to support the development, control, and data reduction of wet-lab experiments. In addition to the BioSPICE software products, a project Web site supports information exchange and community building.