Morning Session I: The Genome

1. Mitomaster: analytical tool for mtDNA.
   Martin Brandon, Pierre Baldi, Douglas C. Wallace

   The recent surge in molecular biological information is increasing our understanding of the underlying mechanisms of genetic diseases; however, the volume and complexity of this information makes it difficult for the average clinician to incorporate into a diagnostic framework. Databases that catalog known disease-causing mutations are one way of augmenting diagnostic abilities, but their usefulness is quite limited due to the relatively small proportion of mutations that have been characterized. A more effective strategy would combine these catalogs with information about the molecular mechanisms involved, so that the effect of novel mutations might be predicted by extrapolation. In fact, this is the approach commonly used by researchers in the lab, but it is not practical in a general clinical setting, where the diagnostician lacks such specialized knowledge. Mitomaster is a system being developed that is able to computationally analyzed the variation within mtDNA and provide links to relevant clinical publications.

2. Evolutionary rate and genome size differences between Arabidopsis lyrata and Arabidopsis thaliana.
   Leah J. DeRose-Wilson, Michael Sweredoski, Pierre Baldi, Brandon Gaut

   Arabidopsis lyrata is an outcrossing relative (separated by ~5MY) of Arabidopsis thaliana. Outcrossing species are expected to have a greater effective population size due to their greater effective recombination rate than their selfing relatives. This expectation means that natural selection is more efficient in an outcrossing species than in an inbreeding relative. Breeding system changes between this two species predict that there would be genome wide differences in evolutionary rate between these species. Additionally, A. lyrata has a genome that is 1.5X greater than that of A. thaliana. It is thought that this increase in genome size is due to different rates of insertion/deletion in non-coding DNA. There is some evidence that introns are consistently smaller in A. thaliana—but this is not nearly enough to account for the observed genome size differences. To calculate an evolutionary rate for the A. lyrata genome, I have sequenced ~100 transposable-element-containing intergenic regions, an assumed neutrally evolving portion of the genome, in A. lyrata. These data will also be useful for investigating the role of transposable elements and intergenic DNA insertions and deletions in the observed size differences between A. lyrata and A. thaliana genomes.

3. Assessing genomic and phylogenetic factors contributing to Basho transposon population frequencies in Arabidopsis thaliana.
   Jesse Hollister, Dennis Kibler, Brandon Gaut

   Transposable elements (TEs) make an important contribution to the size, structure, and variability of plant genomes. Roughly one quarter of the TEs in the Arabidopsis thaliana genome consist of a group of Helitron TEs known as Basho. Helitrons are a recently discovered group of DNA TEs that
replicate via a rolling-circle mechanism. In this study, we constructed phylogenetic trees from 441 Basho sequences extracted from the published A. thaliana genome. We determined the population frequencies of 278 these loci in a panel of 48 A. thaliana individuals. The influence of element age (approximated by branch length from the phylogenies), element sequence length, gene density, and recombination rate on population frequency for all Basho loci were assessed. The relative importance of these factors in the population dynamics, fixation, and deletion of Basho elements will be discussed.

4. Methionine biosynthesis limits adaptation to high temperature in E. coli.
   Nancy M. Aguilar, Jason Schlumbohm, Albert F. Bennett, Anthony D. Long

   In Escherichia coli, degradation of the first enzyme in the methionine biosynthesis pathway may be a critical limitation to growth at high temperatures. To determine if this pathway was subject to change by natural selection, we measured fitness and growth rate in 6 lines of E. coli that had evolved for 2000 generations at 41.5°C. In 5 of the 6 lines, stabilization of the methionine pathway was found to be an important component of their increased fitness relative to their ancestral strain. To determine the basis for adaptation in this pathway, the 8 genes of the biosynthesis pathway were sequenced from the ancestors and evolved lines. The significance of mutations found in this pathway was difficult to ascertain without any structural data for the proteins. Thus, comparative microbial genomics was used to identify regions that may be targets of natural selection. Based on this approach, mutations from the experimentally evolved lines do not appear to be of functional significance, however, this has yet to be confirmed in vivo with site-directed mutagenesis.

5. Reliable large scale multiplex PCR.
   Christopher D. Wassman, Stuart J. Macdonald, Anthony D. Long, Richard H. Lathrop

   Polymerase chain reaction (PCR) provides a method to amplify a specific segment of DNA. Multiplex PCR provides a method to amplify multiple DNA segments simultaneously, using multiple primer pairs in the same reaction. Multiplex PCR is a foundational technology enabling detection of pathogens and other microorganisms, whole-genome sequencing, forensic sequence analysis, and high throughput genotyping. Reliable multiplex PCR offers significant reduction of both time and monetary costs. In practice, however, large scale multiplex PCR (more than 20 amplicons) often requires cumbersome and costly experimental troubleshooting with no guarantee of success. Therefore, improved methodology is demanded for large scale applications. Here we attempt to improve the success of large scale multiple PCR via rigorous and efficient computational primer selection. This is accomplished by selecting multiplex sets of primer pairs to minimize predicted melting temperatures of undesirable annealing events. Melting temperatures are predicted via highly accurate nearest neighbor thermodynamic parameters sets. Results from various minimization algorithms are examined, and selected primer sets are compared to primer sets generated via existing non-multiplex PCR design tools.

   Eric T. Wang, Pierre Baldi, Robert K. Moyzis

   Using the > 3.9 million single-nucleotide polymorphism (SNP) genotype datasets from Perlegen Sciences (Hinds et al., Science 307, 1072-1079, 2005), and the updated International Human HapMap Project (Nature 437, 1299-1320, 2005), a probabilistic search for the landscape exhibited by positive Darwinian selection was conducted. By sorting each high frequency allele by homozygosity, we search for the expected decay of adjacent SNP linkage disequilibrium (LD) at recently selected alleles, eliminating the need for inferring haplotype. We designate this approach the LD decay (LDD) test.
(Wang et al., PNAS 103, 135-140, 2006). The method relies only on high-heterozygousity SNPs for analysis, thereby avoiding the problems associated with applying traditional population genetics tests for selection to the high ascertainment bias Perlegen and HapMap data sets. Using this sensitive LDD test, the fingerprint of recent inferred selection was uncovered for over 100 times more loci (~ 5,000) than other published global approaches using less sensitive tests. Even with this increased sensitivity, simulation studies indicate that the LDD test, at the Mb scale employed, effectively distinguishes selection from other causes of extensive LD, such as inversions, population bottlenecks and admixture. The genes identified by the LDD test were clustered according to Gene-Ontology (GO) categories. Based on over-representation analysis, several predominant biological themes are common in these selected alleles, including host-pathogen interactions, reproduction, DNA metabolism/cell cycle, protein metabolism, and neuronal function. While the number of selected alleles identified is large, it is similar to the estimated number obtained for artificial selection (by humans) on the maize genome (Wright et al., Science 308, 1310-1314, 2005). We have calculated allele ages for all selected sites, indicating that most of the human selective events occurred in the last 10,000-40,000 years, a time of major population expansion out of Africa followed by regional shifts from hunter-gatherer to agrarian societies. The acceleration of these adaptive events in the last 10,000 years suggests that gene-culture interactions directly or indirectly shaped the human genome.

**Morning Session II: The Transcriptome**

1. **Deciphering the regulatory code controlling hair follicle cycling.**
   Kevin K. Lin, Darya Chudova, Alexander T. Ihler, Padhraic Smyth, Bogi Andersen

   The hair growth cycle is an example of a cyclic process that is well characterized morphologically but incompletely understood at the molecular level. Time-course profiling of all components of the skin is needed to comprehensively characterize the complex molecular mechanisms involved in hair follicle morphogenesis and cycling, which involves both the epithelial and mesenchymal compartments. We previously reported the identification of hair cycle-associated genes from time-course gene expression profile data of the complex skin tissue by using a computational approach. However, some of these genes may only be playing a key role in hair follicle morphogenesis and not necessarily involved in hair follicle cycling. Hence, we refined this list of hair-cycle associated genes by performing microarray analysis on the second hair growth cycle, as well as the depilation-induced hair cycle, and combining it with the available dataset of the first hair cycle. Probabilistic models were built to integrate microarray data of three hair growth cycles and identify hair cycle-associated genes, as well as genes involved in hair follicle morphogenesis and skin injury/wound healing. Hair cycle-associated genes are then clustered using a mixture model algorithm we developed for time-course gene expression profile data with replicated measurements. To decipher the regulatory code controlling hair follicle cycling, we predicted functional transcription factor binding sites by incorporating comparative genomics and over-representation of sites in the upstream regions of hair cycle-associated genes that are co-expressed both temporally and spatially.

2. **Characterization of RAR target genes.**
   Benjamin Huang, S. Joshua Swamidass, Pierre Baldi, Bruce Blumberg

   Retinoic acid receptors (RARs) belong to a superfamily of proteins called nuclear receptors. RARs have been demonstrated to play critical roles in a large number of biological processes, including anteroposterior patterning. In order to better understand this complex signaling pathway during early development, micro-array analysis was employed to identify *X. laevis* genes that were upregulated or downregulated by modulating RAR signaling. From a subset of these genes, quantitative real time PCR
tests were performed to identify direct and indirect RAR targets. By using modeling tools, such as Footprinter and Phylogibbs, we wish to identify shared motifs from these genes and then use the presence or absence of these motifs as predictors of direct and indirect targets. Once identified we wish to clone the elements into reporter gene constructs to test in vivo their ability to respond to RAR signaling.

3. **In silico prediction and functional identification of σ28-regulated genes.**
   Hilda H. Yu, Dennis Kibler, Ming Tan

   One mechanism of gene regulation in the human pathogen, *Chlamydia*, is through the use of alternative forms of RNA polymerase which direct the transcription of specific classes of genes. We have previously demonstrated that σ28 RNA polymerase is active and can transcribe *hctB*, a gene expressed late in the developmental cycle when metabolically active forms of chlamydiae are being converted to the more compact, infectious forms. Our current goal is to identify other genes that are transcribed by σ28 RNA polymerase using bioinformatics methods. In collaboration with Dr. Dennis Kibler, we have developed a computer algorithm, using two different probability weight matrices, one based on known σ28 promoters in other bacteria, and the other based on functional data obtained by assaying the effect of each single base substitution in the *hctB* promoter on transcriptional activity in vitro. We selected 16 predicted σ28 promoters to test with our in vitro transcription system, and confirmed σ28-specific promoter activity for 5. Among these 5 new σ28-regulated genes is *dnaK*, which has been shown to be upregulated more than 10-fold under heat shock. This discovery suggests a potential, more encompassing, role for σ28 RNA polymerase in regulating the general stress response in *Chlamydia*.

4. **Identification of novel transcriptional regulators in Chlamydia.**
   Johnny Akers, Bob Chan, Richard Lathrop, Dennis Kibler, Ming Tan

   The pathogenic bacterium *Chlamydia* is an obligate intracellular parasite whose genes are coordinately expressed during its developmental cycle. Although there is evidence of transcriptional regulation in *Chlamydia*, the mechanisms have not been well defined. In particular, very few transcription factors have been predicted, and only a handful have been studied for function. We have developed an approach utilizing both bioinformatics methods and functional testing to study novel transcriptional regulators in *Chlamydia*. Using a structure prediction Meta Server from BioInfo.PL, we searched the chlamydial genome for proteins that contain structure domains that are characteristic of transcription factors. A putative transcription factor, TroR, was identified. TroR is not present in the annotated chlamydial genome, and we predict that TroR regulates the expression of the divalent cation transport genes in *Chlamydia* based on comparative genomics. We have also developed an approach to identify unrecognized transcription factors in *Chlamydia* by first identifying cis-acting DNA regulatory elements that serve as binding sites for the transcription factors. The algorithm searches the chlamydial genomes for sequences containing characteristics of transcription factor binding sites such as inverted or direct DNA repeats. We hypothesize that conserved DNA motifs located in intergenic regions are likely to serve as binding sites for transcription factors. We will use selected DNA motifs as the ‘bait’ for directly purifying candidate transcription factors on the basis of specific binding to each DNA sequence. These DNA binding proteins will be further tested in vitro with functional assays to determine if they are transcription factors that can regulate promoter activity.
5. **Pervasive genetic variation for alternative splicing decisions and cis-regulation of gene expression in *Drosophila melanogaster***.
   Jonathan Gruber, Klemens J. Hertel, Brandon S. Gaut, Anthony D. Long

We have performed an experiment that demonstrates segregating genetic variation in *Drosophila melanogaster* for two critical gene expression phenotypes: alternative splicing and cis-regulation. This represents a key step to understanding the genetic architecture of gene expression. If a large fraction of genes has detectable levels of regulatory variation, are regulatory alleles rare in frequency and of large effect, or at intermediate frequency and of less dramatic effect? The answers to these questions are critical to addressing the often-postulated idea that regulatory change is more key to evolution than is change at the protein-coding level. We have begun to address these questions by focusing on a small set (20-40) of genes, for which we quantified relative expression in up to 15 genetic backgrounds with high levels of biological and technical replication. We crossed 15 highly inbred “tester” strains to two “reference” strains, generating heterozygous progeny for all genetic differences between the tester and the reference parents. We then extracted RNA from single flies in a randomized, blocked design (15-20 progeny per cross, more than 640 RNA extractions in all). After synthesizing cDNA, we performed high-throughput “phenotyping,” examining splicing and relative allelic expression using an Oligo Ligation Assay (OLA). When crossed to the same reference strain, significant differences among tester stains in the expression ratio of two SNP alleles imply the existence of cis-regulatory variants. Variation in alternative splicing choices is indicated when a splice-junction assay yields significantly different ratios among strains. Our method has high power to detect variation of small effect. Regulatory alleles, when detected, are often observed to segregate at intermediate frequencies.

*Afternoon Session I: The Proteome*

1. **New backbone assignment strategies based on precision amino acid labeling.**
   Kevin Donovan, Mike Sweredoski, Pierre Baldi, A.J. Shaka

The HSQC and HNCA spectra of calmodulin were simulated for precision labeled samples where only specified residue types were isotopically labeled. There were four precision labeled samples, with 4 or 5 labeled residue types in each sample. The labeling schedule was generated from an optimization strategy that minimizes the number of predicted overlaps between peaks. The resulting spectra were compared to corresponding spectra from uniformly labeled samples. The comparisons showed a dramatic increase in spectral resolution, and ease of assignment and thus offers a strong advantage in assigning crowded multidimensional protein NMR spectra.

2. **Choosing where to look next: applying active learning to choose informative p53 cancer rescue mutants.**
   Samuel A. Danziger, Jue Zeng, Rainer K. Brachmann, Richard H. Lathrop

Many branches of biological science seek to reduce the time and expense incurred by *in vitro* experimentation by using computer models. Current computer models are frequently incomplete or inaccurate. Machine learning techniques may be used to compensate for defects in these models. Furthermore, Active Learning techniques for choosing which expensive data point to acquire next make it possible for biologists and computer scientists to work together to rapidly build the best possible models. To demonstrate this, five Active Learning techniques are used simultaneously to predict a set previously unknown p53 mutants. The best of these techniques improved a baseline 56% accuracy without Active Learning to 77% accuracy with Active Learning.
3. **Pinot Noir: protein informatics for computational NMR optimizers.**
   Michael Sweredoski, A.J. Shaka, Pierre Baldi

   Strategic NMR labelling schemes and NMR backbone assignment techniques rely on accurate models of the distribution of the backbone chemical shifts. By simultaneously mining the Biological Magnetic Resonance Data Bank (BMRB) and Protein Data Bank (PDB), we are able to create a mapping between the two databases and build models which are more accurate than those previously published. This mapping allows us and other researchers to study the relationship between the backbone chemical shifts and the structural features of proteins. Our mapping and statistics, which we curate from over 2,000 proteins, incorporate protein structural features such as secondary structure and solvent accessibility. As data is deposited in these databases, we continuously update the statistics and figures on our web site at http://www.ics.uci.edu/~baldig/PDB2BMRB/.

4. **Investigations of JNK substrate binding motifs via computational and experimental studies.**
   Thomas Whisenant, Ryan W. Benz, David Ho, Pierre Baldi, Lee Bardwell

   The JNK pathway as part of mitogen activated protein kinase signaling plays a critical role in important cellular processes such as apoptosis and transformation, as well as diseases including diabetes, muscular dystrophy and Parkinson’s disease. However, to date few JNK substrates have been identified complicating additional studies of these pathways. Furthermore, experimental methods used to study JNK substrates can be quite time consuming, suggesting the use of computational methods to help direct the experimental efforts. In this work, Hidden Markov Models (HMMs) have been used to search for a characteristic JNK binding motif in a database of protein sequences in order to identify potentially novel JNK substrates. Based upon the HMM results, the top scoring sequences were then experimentally tested for JNK binding using a peptide blot. Initial results from the peptide blot show that the HMM, despite the limited training set, is able to find JNK binding motifs from full length protein sequences. Further analysis of the HMM and peptide blot, along with the implications for finding novel JNK substrates will be presented.

5. **Cyclic modular β-sheets.**
   R. Jeremy Woods, Omid Khakshoor, Justin O. Brower, Mehrnoosh Hashemzadeh, Wade A. Russo, James S. Nowick

   This paper presents a new class of modular macrocyclic β-sheets that adopt robust, well-defined β-sheet structures in aqueous solution and can be linked to form multivalent structures. These cyclic modular β-sheets, which include the unnatural amino acid Hao as a β-strand template and two δ-linked ornithine β-turn mimics, are easily and quickly prepared by standard peptide synthesis techniques. Spectroscopic and computational studies with cyclic modular β-sheets containing several different sequences demonstrate the largely non-sequence dependent β-sheet folding of these 42-membered ring macrocycles and provide some insight into the strengths and weaknesses of the design.

**Afternoon Session II: Molecular Modeling**

1. **Tertiary structure prediction with predicted structural features and fragment assembly.**
   Arlo Z. Randall, Jianlin Cheng, Mike Sweredoski, Luis Villarreal, Pierre Baldi

   Our approach to ab initio tertiary structure prediction combines the use of predicted structural features, a fragment library, physical atomic properties, and energy terms derived from PDB statistics.
The structural features used are secondary structure, relative solvent accessibility, and a residue level contact map at a distance cut-off of 12 Å. The predicted structural features are used in the energy function. We use a database of protein fragments of length nine, constructed from a redundancy reduced set of PDB structures. Our method performed well against other automated methods on hard targets (NF and FR/A CASP designations) in CASP 6.

2. **Computational identification of a new tuberculosis drug leads.**
   S. Joshua Swamidass, Sheryl Tsai, Pierre Baldi

   The “Great White Death”, tuberculosis (TB) remains one of the greatest global health challenges. The increasing prevalence of MDR TR motivates the development of new pharmaceuticals in order to reliably cure TB infections. The waxy cell coat is the key. This coat prevents most antibiotics from entering the bacterium. Some of the antibiotics that do manage to enter the cell inhibit synthesis of mycolic acid, a key component of the cell coat. This inhibition reliably kills TB cells. So the coat is both a pharmaceutical barrier and a drug target. Our collaborator crystallized the AccD5 enzyme, critical for the synthesis of key fatty acids. Using this structure and a combination of similarity and docking based filters, we proposed a set of small molecules inhibitors of AccD5. Laboratory testing confirmed two of them as in vitro inhibitors of AccD5 at ~10 micro molar concentration. Further cell culture testing confirmed that one of the leads inhibits mycobacterium at ~50 micro molar. Further computational studies are being performed to further refine these hits into better leads.

3. **Combinatorial reactions: applications and discovery in chemical informatics.**
   Jonathan H. Chen, S. Joshua Swamidass, Peter Phung, Pierre Baldi

   Small molecules play a fundamental role in organic chemistry and biology. They can be used as combinatorial building blocks for chemical synthesis, as molecular probes in chemical genomics and systems biology, and for the screening and discovery of new drugs and other useful compounds. The development of chemoinformatics, the study of the information space of such molecules, has been hampered by the lack of large, publicly available, comprehensive repositories of molecules. By comparison, the two main driving forces behind the bioinformatics expansion have been the development of high throughput methods and the corresponding public availability of large repositories (GenBank, Swissprot, PDB, etc) and the development of search algorithms (BLAST) and related statistical machine learning techniques to analyze the data. ChemDB, a public database of small molecules available and searchable over the Web was developed to address this gap. ChemDB is built using the digital catalogs of over a hundred vendors and other public sources and is annotated with information derived from these sources as well as from computational methods, such as predicted solubility and 3D structure. The current version of the database contains approximately 4.1 M commercially available compounds. Considering virtual compounds that can be synthesized from building blocks in the ChemDB can expand the repository’s size further. Annotating functional groups and applying *in silico* reactions allows one to explore the virtual space of compounds indirectly available by applying reactions to the components in ChemDB. Such reaction schemes can be applied to other novel usages such as screening for polymer candidates and interactive education tutorials. These “knowledge-based” approaches generally require pre-specification of the reaction schemes however, so we have also generated our own reaction schemes based on general, pseudo-mechanistic principles in an effort to discover new reaction schemes by systematic search that may have been overlooked by chemist intuition.
4. **Computer simulations of a voltage sensor in a lipid bilayer.**
   Alfredo Freites, Doug Tobias, Stephen White

Voltage-dependent ion channels open and close in response to changes in transmembrane (TM) potential as a result of the motion of voltage-sensor domains (VSD). We have performed a series of molecular dynamics (MD) simulations aimed at understanding the stability and function of the VSD of voltage-dependent potassium (Kv) channels. To this end, we have modeled the following systems in a phospholipid bilayer with excess water: the unit of electric charge in VSD as the guanidinium ion; the transmembrane helix responsible for the voltage-sensing, the so-called S4 helix; and the whole VSD of the KvAP, a Kv channel from an archaeabacterium. Our results reveal that the fluid nature of the lipid membrane provides the necessary hydration to these charged systems allowing their stable integration in the lipid membrane. Based on this principle, it is possible to reconcile a large collection of biochemical and biophysical studies of Kv channels that, according to their classical interpretation, were at odds with the more recent direct structural evidence.

5. **DNA deformation energy as an indirect recognition mechanism in protein-DNA interactions.**
   Kimberly A. Aeling, Nicholas R. Steffen, G. Wesley Hatfield, Richard H. Lathrop, Donald F. Senear

Proteins that bind to specific locations in genomic DNA control many basic cellular functions. Proteins detect their binding sites using both direct and indirect recognition mechanisms. Deformation energy, which models the energy required to bend DNA from its native shape to its shape when bound to a protein, has been shown to be an indirect recognition mechanism for one particular protein, Integration Host Factor (IHF). This work extends the analysis of deformation to two other DNA-binding proteins, CRP and SRF, and two endonucleases, I-CreI and I-Ppol. Known binding sites for all five proteins showed statistically significant differences in mean deformation energy as compared to random sequences. Binding sites for the three DNA-binding proteins and one of the endonucleases had mean deformation energies lower than random sequences. Binding sites for I-PpolI had mean deformation energy higher than random sequences. Classifiers that were trained using the deformation energy at each base pair step showed good cross-validated accuracy when classifying unseen sequences as binders or non-binders. These results support DNA deformation energy as an indirect recognition mechanism across a wider range of DNA-binding proteins. Deformation energy may also have a predictive capacity for the underlying catalytic mechanism of DNA-binding enzymes.

Afternoon Session III: Systems Biology

1. **Systems biology modeling of mitochondria OXPHO pathway and ROS generation.**
   Ivan Y. Chang, Pierre Baldi, Doug Wallace

Mitochondria are the power house of eukaryotic cells, and are also the major sites of Reactive Oxygen Species (ROS) generation. Recently, there has been a growing interest in the study of cellular energetic deficiency, and its relationship with major age related diseases such as diabetes, Alzheimer’s, ALS, etc. Current models of mitochondria oxidative phosphorylation (OXPHO) do not account for the ROS production and thus are not capable to model aging through the accumulation of ROS induced damages. This project is an extension of the OXPHO model with several path of ROS production. The framework is based on the principle of mass balance in which a system of differential equations governing the concentration of metabolites is generated from the balance between in-fluxes and out-fluxes of the metabolites through components in the pathway. Each component is represented by a flux equation that describes its substrates and products turnover rate, and its formulation can be based on
empirical studies or first principle. Once the parameters of the flux equations are optimized, the set of differential equations are then solved through Mathematica. The results are then compared with experimental results to test for the validity of the model.

   Tarek S. Najdi, Chin-Rang Yang, Bruce E. Shapiro, Eric D. Mjolsness, G. Wesley Hatfield

In our effort to elucidate the systems biology of the model organism, *Escherichia coli*, we have developed a mathematical model that simulates the allosteric regulation for threonine biosynthesis pathway starting from aspartate. To achieve this goal, we used kMech, a Cellerator language extension that describes enzyme mechanisms for the mathematical modeling of metabolic pathways. These mechanisms are converted by Cellerator into ordinary differential equations (ODEs) solvable by Mathematica™. In this paper, we describe a more flexible model in Cellerator, which generalizes the Monod, Wyman, Changeux (MWC) model for enzyme allosteric regulation to allow for multiple substrate, activator and inhibitor binding sites. Furthermore, we have developed a model that describes the behavior of the bifunctional allosteric enzyme aspartate kinase I-homoserine dehydrogenase I (AKI-HDHI). This model predicts the partition of enzyme activities in the steady state which paves the way for a more generalized prediction of the behavior of bifunctional enzymes.

   Alex Sadovsky, Tigran Bacarian, Ana Campilho, Marcus Heisler, Elliot Meyerowitz, Ben Scheres, Eric Mjolsness, Pierre Baldi

Interaction between mitotic activity and growth kinematics is a central question in plant development and physiology. Tissue movement and formation of new cell walls affect such vital aspects of a plant’s functionality as genetic-, metabolic- and biomechanical regulation. Departing from the Mitotic Behavior Hypothesis that a cell undergoes mitosis upon reaching a threshold size, we formulate a mathematical model to study the relation between mitotic activity and growth kinematics in the stelar region of the *Arabidopsis thaliana* root meristem. By averaging across the transverse sections and parametrizing all biologically relevant quantities by height level, we idealize the root stele as a 1-dimensional segment. To provide input data for the model, measurements are collected from a root specimen by time-lapse imaging and are analyzed using image-processing algorithms. As an outcome of the experiment, two adjacent subsegments are identified by a sharp transition in the mitotic rate at their common endpoint. From the mitotic rates observed in each subsegment, our model predicts the functional form of the corresponding velocity profiles, determined up to a choice of a coefficient. The specific velocity profiles are obtained by fitting the coefficient to the experimental data on the velocity. From the mitotic activity- and velocity profiles, our model provides a simple way of calculating the average mitotic threshold cell length in each subsegment. In addition, the model yields an equation for calculating the cell cycle duration knowing only the velocity profile. All equations in the model are \{em autonomous\}, hence independent of the specific cellular configuration in the tissue. These results show that, in a longitudinal subsegment of the stele, the ratio of the mitotic rate to the strain rate is constant to a good approximation, and that the mitotic threshold cell length and the cell cycle duration can be calculated from the mitotic rate- and velocity profile in the subsegment. This provides evidentiary support of the Mitotic Behavior Hypothesis.
   Michael Duff, Darya Chudova, Barbara Wold, Eric Mjolsness, Padhraic Smyth

The reconstruction of genetic networks has been called the “Holy Grail” of functional genomics. The core task is to use gene expression array data to identify the causal structure of a gene network---to determine, for each designated gene in the network, the genes that are directly influenced by that gene. Many approaches have been advanced for this problem of system identification, including Bayesian methods of inference based upon static or dynamic network models with discretized expression-level values. Fundamental theoretical issues of network identifiability must be addressed, and there are practical limits to what can be achieved, particularly with regard to sparse and noisy data. Our efforts maintain a Bayesian framework, but consider a generalized class of flexible stochastic nonlinear dynamical systems for modeling gene expression time-series data. These dynamical models are combined with principled Bayesian methods for inferring causal structure in the underlying unknown genetic circuit. Theoretical aspects include the adoption of marginal likelihood and approximations to marginal likelihood as network structure scoring criteria that embody Occam’s razor. Practical issues include robust optimization over very-large parameter spaces, and the application of efficient Monte Carlo procedures (ranging from Gibbs and Metropolis samplers to reversible-jump and parallel-tempering schemes) for sampling from the Bayesian posterior distribution over network structures.