

SYSTEMS BIOLOGY CENTER
NEW YORK

2014 SUMMER RESEARCH PROGRAM IN SYSTEMS BIOLOGY

POSTER SESSION

DATE:
FRIDAY AUGUST 8, 2014

LOCATION:

ICAHN BUILDING 12 Floor
(Outside of room 12-59A)

1425 MADISON AVE, NEW YORK, NY 10029
ICAHN SCHOOL OF MEDICINE AT MOUNT SINAI

TIME:
1:00PM - 2:00PM



Icahn
School of
Medicine at
Mount
Sinai



NIGMS

For more information about the program
please visit www.sbcny.org/2014_program.htm



SBCNY is supported by Grant Number P50GM071558
from the National Institute of General
Medical Sciences

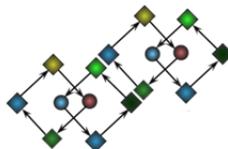
Structure-based Ligand Discovery for SLX4 by Virtual Screening

Henna Ahmed^{1, 2}, Avner Schlessinger PhD²

¹Macaulay Honors College at CUNY Hunter College

²Department of Pharmacology and Systems Therapeutics, Systems Biology Center New York, Mount Sinai School of Medicine, New York, NY 10029

Glioblastoma multiforme (GBM) is the most common and lethal type of brain cancer, with median survival rate of 12-15 months. The DNA repair mechanism plays a large role in the resistance of GBM to treatments like chemotherapy and radiotherapy. An important aspect of the DNA repair pathway is the repair of interstrand crosslinks (ICLs) to allow the progression of the cell cycle. SLX4 is a tumor suppressor that aids in this repair. A deficiency in SLX4 results in sensitivity to DNA crosslinking agents which leads to ICLs. Consequently, the absence of SLX4 results in interference with the cell's DNA repair mechanism and thus, inhibition of the progression of the cell cycle. Furthermore, using systems-biology approach the Iyengar lab has identified SLX4 as a potential target for developing drugs against GBM. Our goal is to identify inhibitors for SLX4 using structure-based discovery approach. Although the full structure of SLX4 is unknown, a crystal structure of the SLX4 region that binds TRF2 was recently determined. Our goal was to find ligands that inhibit the binding of SLX4 in TRFH region of TRF2. In essence, we hypothesize that this inhibition would decrease the proliferation of glioma cells. Using virtual screening of 1,748,064 compounds from the ZINC library we identified 26 potential inhibitors for the SLX4/TRFH interactions, which will be tested experimentally by our collaborators. This research is expected to provide valuable information regarding the SLX4-TRF2 interaction and its role in SLX4 depletion.



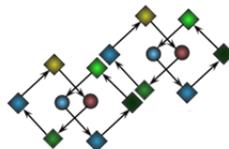
Targeted Drug Combinations for Glioblastoma Multiforme Through Computational Modeling

Matthew DiStefano^{1,2}, Mehdi Bouhaddou², Marc Birtwistle²

¹Oceanside High School, ²Department of Pharmacology and Systems Therapeutics, Systems Biology Center New York, Icahn School of Medicine at Mount Sinai, New York, NY 10029

Glioblastoma Multiforme (GBM) accounts for 15 percent of all brain tumors and its ability to invade different portions of the brain makes the complete surgical removal of tumors nearly impossible. Since GBM relies on several different cellular signaling pathways to proliferate, the exact configuration of which depends largely on the particular mutations the tumor possesses, a more accurate way of developing personalized drug combination therapies and treatments based on a tumor's genomic makeup should lead to higher treatment success rates.

To achieve this goal, a kinetic computational model was built based on mass action kinetics and using ordinary differential equations, which can simulate the interactions between ~800 proteins in silico. In order to test and refine the accuracy of our model we are using data from the Cancer Cell Line Encyclopedia (CCLE) by the Broad Institute, the Cancer Genome Project (CPG) by the Wellcome Trust, and The Cancer Genome Atlas (TCGA) to obtain genomic, mutational, and pharmacological data to be used as inputs for our model. These three data sources give us the ability to tailor the model, validate its predictive power, and analyze results for potentially useful drug combinations, respectively.



Hepatocellular Carcinoma Patient Classification with Enrichment Vectors

Axel S. Feldmann^{1,2}, Yan Kou², Yujin Hoshida³, Avi Ma'ayan²

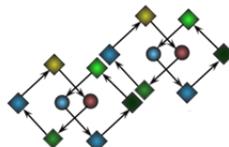
¹Hunter College High School, 71 E 94th St, New York, NY 10128

²Department of Pharmacology and Systems Therapeutics, Systems Biology Center New York, ³Department of Medicine, Icahn School of Medicine at Mount Sinai, One Gustave L. Levy Place Box 1215, New York, NY 10029

Hepatocellular carcinoma is one of the deadliest cancers accounting for over 650,000 deaths worldwide each year. Prognosis of the disease is poor with only 20% of patients surviving one year after diagnosis and only 5% surviving through five years. Only 10% of patients are diagnosed early enough for the cancer to become operable. Hence, it is essential to improve patient subtype classification and develop novel therapeutics. Unsupervised clustering of patients in an effort to identify distinct prognosis groups has been done by examining genome-wide mRNA expression levels profiled with microarrays [1, 2]. However this approach only identified moderate differences in patient outcomes within clusters cases [1]. Unsupervised clustering of patients can also identify sets of genes that explain molecular mechanisms besides providing prognosis predictions [2]. However, such approaches are limited by the fact that they treat all genes in a similar manner instead of highlighting genes that are active in pathways and/or mechanisms that may have a greater impact on prognosis. Here we present an alternative unsupervised clustering approach that converts gene expression data from individual patients to enrichment vectors. Such enrichment vectors are created by comparing differentially expressed genes in individual patients with gene set libraries containing gene sets associated with various biological processes as well as drug treatments. After clustering patients based on the enrichment vectors, Kaplan Meier survival curves are used to identify clusters of patients with significantly better or worse survival outcomes. We identified unique sets of genes common in both poor and favorable outcome clusters and associated these genes with their molecular pathways. We also identified FDA-approved drugs that can potentially improve prognosis for individual patients by reversing the gene expression by down regulating up-regulated genes and up regulating down-regulated genes. Our analysis and methods can potentially improve our understanding of liver cancer molecular mechanisms, as well as suggest novel treatment options for individual patients. The methods used to generate the results are open source and can be applied to study other cancers.

[1] Hoshida, Yujin, et al. "Gene expression in fixed tissues and outcome in hepatocellular carcinoma." *New England Journal of Medicine* 359.19 (2008): 1995-2004.

[2] Lee, Ju-Seog, et al. "Classification and prediction of survival in hepatocellular carcinoma by gene expression profiling." *Hepatology* 40.3 (2004): 667-676.



Orders of Magnitude Computational Acceleration for Large Scale Cardiac Simulations

Farah Fouladi^{1,2}, Eric Sobie²

¹Colgate University, Hamilton NY,

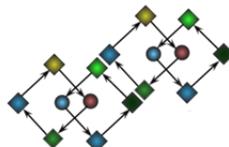
²Department of Pharmacology and Systems Therapeutics, Systems Biology Center New York. Icahn School of Medicine at Mount Sinai, New York, NY 10029

Through mathematical modeling we can predict the effect of changes in parameters that are difficult to measure experimentally. This requires multiple simulations of the model for every parameter variation. While this type of analysis is feasible for isolated cells, running many simulations of large systems of cells is extremely time consuming. To decrease the execution time of these mathematical models, I have developed a parallelized simulation environment, which utilizes the architecture of a graphics processing unit.

The GPU has many processor cores and the ability for thousands of threads to run concurrently on those cores. For my research, a mathematical model of the human ventricular myocyte (ten Tusscher & Panfilov, *Am J Physiol* 291: H1088–H1100, 2006) was programmed in CUDA, a programming language for graphics processors. Taking advantage of the parallel hardware, cell computations are completed on the GPU by many threads running simultaneously. Additionally, different simulations of the model are executed in parallel. To validate this computational design, cell membrane voltage values were compared with an already existing model in MATLAB. Results showed that the added parallelization has no effect on the computational aspect of the model and that the execution time of the CUDA program is orders of magnitude quicker than that of the MATLAB program.

Using this novel software, I analyzed one cause of reentry in cardiac myocytes. Reentry occurs when an electric propagation loops back on itself, abnormally re-exciting cells. There is a short window of time during which a stimulus can excite cardiac tissue and cause a reentry effect due to refractory tissue blocking action potential propagation in only one direction. My program is able to efficiently identify the stimulus-timing interval when reentry occurs in a loop of 1,000 cells.

This parallelized simulation environment minimizes computational execution time and provides a framework for further analysis of more complex and physiologically relevant systems of cells.



Subcellular process enrichment analysis of high-throughput data identifies underlying biology

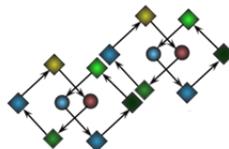
Manhin Lam^{1, 2}, Jens Hansen², Ravi Iyengar²

¹City University of New York - New York City College of Technology, ²Department of Pharmacology and Systems Therapeutics, System Biology Center New York. Mount Sinai School of Medicine

Cells have a variety of different functions or sub-cellular processes. Depending on the cell state, different sub-cellular processes are active. For example, a migrating cell needs sub-cellular processes that are related to a re-organization of the Cytoskeleton, e.g. actin polymerization and de-polymerization, and to vesicular transport (Copper, 2000).

The complex biology causing changes in such sub-cellular processes can be investigated by analyzing high-throughput data. High-throughput technologies such as RNA-Seq and DNA microarrays are powerful experimental tools that enable us to investigate gene expression levels at different conditions and/or time points. In addition, genome-wide siRNA screenings allow to test the influence of many proteins on a certain phenotype, e.g. on the secretory pathway. High-throughput data therefore allows us to identify genes that are involved in certain cellular functions or cellular state changes. However, in order to infer the biological conclusion or insight, a more rigorous statistic is needed. Instead of investigating each gene individually, various ontologies allow us to investigate sub-cellular pathways that are composed of set of genes involved in that function.

In this study, we systematically analyzed published high-throughput datasets referring to the secretory pathway, apoptosis, insulin signaling pathway and the splicing machinery, predicted involved sub-cellular processes by pathway enrichment analysis using different ontologies such as KEGG and Gene Ontology and compared the prediction with the expected biology. We have shown the enrichment analysis results of two of the studies and demonstrate that ontologies are very helpful tools to describe relevant biology.

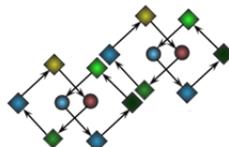


GeneRIFEA: Gene Set Enrichment Analysis by Combining PubMed and GeneRIF

Matthew D. Mottola^{1,3}, Mounica Kamesam^{2,3}, Avi Ma'ayan³

¹Fairleigh Dickinson University, Teaneck, NJ 07666, ²Yorktown High School, Yorktown Heights, NY 10567, ³Department of Pharmacology and Systems Therapeutics, Systems Biology Center New York, Icahn School of Medicine at Mount Sinai, New York, NY 10029 USA

Gene set enrichment analysis (GSEA) is a popular method to analyze genome-wide experimental data in context of prior knowledge. GSEA relies on the existence of gene set libraries that are queried against an input gene list. Gene set libraries are primarily used to organize accumulated knowledge about the function of groups of genes. Each gene set library is made of related gene sets. A gene set includes a list of genes that are associated with a functional term such as a pathway name or a transcription factor that regulates the genes in the set. Gene set libraries can be created by assembling gene sets from diverse contexts. For this project we developed a system that can generate gene set libraries automatically using PubMed and GeneRIF. For this we first created a variety of lists of general search terms covering relevant topics such as: bacteria and viruses, disease, cell-lines and cell-types, cell organelles, tissues, drugs, environmental toxins, of families of genes including: g-coupled protein receptors (GPCRs), ion channels, protein kinases, and transcription factors. Using NCBI's E-utilities, terms are searched in PubMed to return PubMed IDs (PMIDs). The returned PMIDs are then converted to gene symbols using NCBI's Gene Reference into Function (GeneRIF) and genes are counted and sorted for their appearance for each searched term. Fuzzy and crisp gene-set libraries are then constructed for each list of related search term. These gene set libraries are visualized as a set of canvases that arrange related terms based on their gene set content similarity. The whole system is delivered as an enrichment analysis web-based tool. In addition, we utilized these gene set libraries for classification of cancer patients based on gene expression data from the Cancer Genome Atlas (TCGA). This enabled in some cases improved classification of patients based on clinical outcome while providing molecular mechanisms that explain details about each tumor from each individual patient. Overall, the method and system has the advantage of being automatically updatable, but the gene set libraries produced suffer for research focus biases.

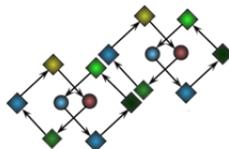


Quantified Self System to Collect and Interpret Longitudinal Urinalysis Data

Matthew Mottola^{1,3}, Axel S. Feldmann^{2,3}, Avi Ma'ayan³

¹Fairleigh Dickinson University, Teaneck, NJ 07666, ²Hunter College High School, 71 E 94th St, New York, NY 10128, ³Department of Pharmacology and Systems Therapeutics, Systems Biology Center New York, Icahn School of Medicine at Mount Sinai, New York, NY 10029 USA

Urinalysis is used to measure and quantify the chemical properties of a urine sample, such as pH and gravity, as well as identify various molecules and cells, such as nitrate, glucose, proteins and various types of blood cells. Urinalysis is central for medical diagnosis for identifying infection and early signs of diseases such as kidney fibrosis and diabetes. A common method to perform urinalysis is through the use of a test strip. Test strips are relatively cheap and once exposed to urine samples can be read against a manufacturer's color scale to obtain semi-quantitative measurements for each indicator. However, strips are read by eye, which is subjective, and the results of the test are often discarded. Automated urinalysis test strip analyzers offer a means of standardizing the reading of test strips. Nevertheless, automated urinalysis test strip analyzers are expensive, limiting their use to mainly medical facilities. For this project we developed a prototype automated test strip analyzer aimed to be cheaper, more accurate, and more suitable for general public home use. Our device uses a microcontroller interfaced with a camera to capture an image of a used test strip within a controlled casing created with a 3D printer. Through the use of a wireless network processor, the captured image of the strip is sent to a server for image processing. The corresponding value of each indicator per sample is logged into an online database and preliminary medical suggestions are sent to the user through e-mail. Trends and in depth analysis can then be performed at the individual level. In the future, we plan to perform the test strip analysis in conjunction with a prototype low cost spectrometer. Comparing the corresponding results obtained from the test strip system will be mapped to the matching urine spectra to develop a urinalysis spectral classifier. This will then allow for a standalone spectrometry based device that will obtain results by analyzing urine with a low cost spectrometer.

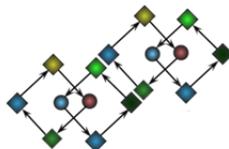


Highly Multiplexed Quantitative Imaging with Combinatorial Fluorescence Probes

Danielle Putur^{1,2}, Marc Birtwistle, PhD²

¹Colgate University, Hamilton, NY 13346, ²Department of Pharmacology and Systems Therapeutics, Systems Biology Center New York, Mount Sinai School of Medicine, New York, NY 10029

There is an increasing need in systems biology for quantitative multiplexing in single live cells over a time course to which fluorescent protein-based technologies may offer a solution. This work is based on the hypothesis that combining combinatorially varied tandem fluorescent protein probes with multi-channel excitation spectral imaging allows for highly multiplexed quantitative imaging. A model has been developed that takes into account reference spectra for each probe, and predicts the amount of each probe present. A smaller-scale project has been executed using seven probes and three excitation channels previously, which found the model is able to correctly predict probe levels. This project aimed to build a large-scale pilot experiment using eighteen probes at four excitation channels by first expressing and characterizing the probes. For each probe, the emission spectra were collected across varied concentrations, and the ideal concentration of the probe was determined. Using that data, we were able to build a reference spectrum for each probe to be used in the model. In the future, we plan to test the model's ability to predict levels of eighteen different probes present in a sample. If successful, this project will lend itself to a highly multiplexed and quantitative imaging technique.

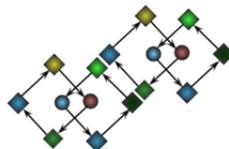


Genome-Scale Identification of Regulatory Networks Through Modular Response Analysis of LINCS L1000 Data

Eric Riesel^{1,2}, Marc Birtwistle²

¹Oceanside High School,²Department of Pharmacology and Systems Therapeutics, Systems Biology Center New York. Mount Sinai School of Medicine, New York, NY 10029

The field of computational biology is becoming increasingly relevant and useful in the technology age. The systems modeling subset of this field has provided biologists with a much deeper understanding of the molecular mechanisms of protein pathways and gene networks and how perturbations affect them. However, in order to model these protein pathways and gene networks, the basic structure of these networks usually have to be experimentally determined. Many programs have been designed for this purpose however they often are imprecise and take an impractical amount of time to run. The development of the theory of Modular Response Analysis (MRA) has revolutionized the field of computational biology by allowing for these connections and values of strengths to be calculated in a smaller time and with much more precision (Santos, 2007). In order to determine network connections MRA is used to calculate connection strength between each gene. When the strength for the connection is positive, gene A promotes gene B, and when the value is negative, gene A inhibits gene B. For each gene added to the network, an experiment perturbing that gene and determining the effects on the expression of all other genes is necessary. Each one of these experiments determines the global response of the network to the disruption of that gene's expression. The global response does not determine direct interaction; instead it only determines the state that the system moves to after knockdown or overexpression of a gene. With each experiment being a column and each gene a row, the collection of the data forms a matrix. The results from hundreds of thousands of gene knockdown and overexpression experiments are available on the Library of Integrated Network-Based Cellular Signatures (LINCS) database. These include the landmark genes, highly non-redundant and expressed genes, and imputed genes (Peck, 2006). This data forms a global response matrix. The point of this study is to determine the network structure and connection strength of the landmark and imputed genes. Determination of new negative and positive feedback loops was also made possible by the network structure and was another aim of this study.



Co-localization of PDE1 and PDE2 determines DA-induced AMPAR trafficking

Philip To^{1,2}, Roy Song², Rosa Tolentino², and Susana Neves²

¹Macaulay Honors College at Hunter, ²Department of Pharmacology and Systems Therapeutics, Systems Biology Center New York Icahn School of Medicine at Mount Sinai, New York, NY 10029

The strength and duration of glutamate synaptic transmission is largely determined by the number of post-synaptic AMPA receptors (AMPA receptors). Modulation of synaptic strength is achieved by changing AMPAR numbers in a process that requires receptor trafficking from intracellular vesicles to the plasma membrane. Trafficking of GluA1-containing AMPARs is regulated by PKA, a cAMP-dependent kinase. Phosphodiesterases (PDEs), enzymes that degrades cAMP, regulate PKA activity and controls dopamine (DA)-induced GluA1 membrane insertion. We have previously shown that the interplay of PDE1 and PDE2 activities regulates AMPAR trafficking. This is due to the allosteric activation of PDE2 by cGMP upon PDE1 inhibition. Moreover, PDE1 regulation of GluA1 trafficking was limited to distal dendritic segments, suggesting selective spatial control. Here we asked whether the localization of PDE1 or PDE2 at the distal dendritic segments regulates GluA1 membrane insertion. Our approach combines live-cell imaging of GluA1 trafficking and computational modeling of signaling to explore the contribution of localization PDE2 to GluA1 trafficking in striatal neurons. We developed a dopamine-dependent model of GluA1 membrane insertion based on Song et al., (2013) and modified the original ODE-model into a partial differential equation based model using an idealized geometry of a neuron with a dendritic diameter of 3 μm . We then tested three different distributions of PDE1 and PDE2: 1) PDE1 and PDE2 are co-localized throughout the dendrite; 2) PDE1 is present only in the distal segment while PDE2 localized throughout the dendrite; or 3) PDE2 is only present in the distal segment and PDE1 is throughout the dendrite. We found localizing PDE1 and PDE2 throughout the dendrite recapitulated best our experimental results. Localizing PDE1 to the distal segment of the dendrite abolished the decrease in GluA1 membrane insertion observed experimentally when inhibiting PDE1. On the other hand, localization of PDE2 to the distal segment of the dendrite led to a small increase in distal GluA1 membrane insertion. To experimentally validate our simulation results, we developed a PDE2 construct that contains the N-terminal localization sequence from PDE2A2 tagged to mCherry (A2A-mCherry). Expression of A2A-mCherry in neurons was limited to the soma and dendritic proximal segments, suggesting that endogenous PDE2A2 expression may have defined spatial range. We found overexpression of A2A-mCherry abolishes the PDE1 inhibition-induced decrease in GluA1 membrane insertion further supporting the cross-regulation occurs between PDE1 and PDE2.

