



**BIO
COM
PLEX
ITY**
INSTITUTE

RESEARCH SYMPOSIUM | 2016

Tuesday, November 1st



Research Symposium | 2016-11-01 | Abstracts

MOLECULAR BASIS OF LIGAND BINDING BY THE ENDOSOMAL ADAPTOR PROTEIN TOM1

Wen Xiong¹, Phillip Choi¹, Xiaolin Zhao¹, Jeff F. Ellena², and Daniel G. S. Capelluto¹

¹Protein Signaling Domains Laboratory, Department of Biological Sciences, Biocomplexity Institute, Virginia Tech, Blacksburg, VA, USA, ²Department of Chemistry, University of Virginia, Charlottesville, VA, USA.

Tom1 (target of Myb 1) plays a role in membrane trafficking by serving as an alternative endosomal sorting complex required for transport (ESCRT)-0 component. Tom1 has been shown to serve as a new phosphatidylinositol 5-phosphate (PI(5)P) effector at signaling endosomes through its VHS domain, delaying cargo degradation in a bacterial infection model. The Tom1 VHS domain also binds ubiquitin moieties in cargo for endosomal transport and degradation; therefore, we hypothesize that the ubiquitin and PI(5)P compete each other for Tom1 VHS binding. In order to address this question, the backbone NMR resonances of Tom1 VHS were assigned. The Tom1 VHS secondary structure prediction scores, using TALOS+, are in good agreement with the secondary structural elements reported for the crystal structure of the protein. Our NMR data revealed that Tom1 VHS interacts with PI(5)P following a fast-exchange regime, with the PI(5)P binding site predicted to be at a region spanning α -helices 6 and 8. In contrast, we found that the ubiquitin-binding site in Tom1 VHS is located at the α -helices 2, 5 and 7 of Tom1 VHS. Despite the binding sites are not overlapped, the ubiquitin and PI(5)P may compete each other by inducing conformational changes in the Tom1 VHS domain upon binding. Also, we identified a conserved central hydrophobic patch at the ubiquitin surface to be the binding site for the Tom1 VHS domain. By providing the molecular basis of the Tom1 interactions, we will generate cargo sorting mechanistic insights, create functionally specific mutations, and precisely manipulate alternative ESCRT-0 proteins.

**STAY
CONNECTED:**

bivt.edu

facebook.com/biocomplexity

twitter.com/biocomplexityvt

youtube.com/c/biocomplexity

google.com/+biocomplexity

Research Symposium | 2016-11-01 | Abstracts

NOVEL OUTBREAK DETECTION METHOD FOR SEXUALLY TRANSMITTED INFECTIONS IN VIRGINIA

Author: Samantha Walmsley | Mentors: Kaja Abbas, Jeff Stover

Objective:

The study objective was to develop a novel statistical method to analyze incidence data of sexually transmitted infections in Virginia to detect aberrations and outbreaks.

Background:

Rapid and accurate detection of outbreaks and aberrations in disease data is vital in helping to maintain healthy populations. This study introduces a novel method of detecting the start of an aberration in incidence data of sexually transmitted infections in Virginia. An aberration is a change in the data itself, whether positive, as in the case of an outbreak, or negative as in a lapse in data reporting.

Methods:

I conducted regression analysis of cumulative incidence data of sexually transmitted infections in Virginia to minimize errors and noise. I fitted two regressions of the cumulative incidence data using a dynamic breakpoint to identify the onset of an outbreak.

Results:

I compared the residuals of the full data set with the pre-aberration regression line, and estimated the disease incidence above the predicted disease burden. The slope of the first fitted regression line of the accumulated cases gives the fitted case rate pre-aberration, while the slope of the second fitted line give the case rate after the aberration. By identifying the breakpoint between the two regression slopes, I estimated the initial date of aberration in incidence of sexually transmitted infections in Virginia and the severity of the aberration.

Conclusion:

I developed and applied a novel statistical method to determine the initial date and severity of aberrations, specifically in the context of sexually transmitted infections in Virginia. . This method can be extended to detect aberrations in incidence of different diseases and geographic regions.

Keywords: aberration, regression, sexually transmitted infections, Virginia, outbreak detection

**STAY
CONNECTED:**

bi.vt.edu

facebook.com/biocomplexity

twitter.com/biocomplexityvt

youtube.com/c/biocomplexity

google.com/+biocomplexity

Research Symposium | 2016-11-01 | Abstracts

STRUCTURAL, THERMODYNAMIC, AND PHOSPHATIDYLINOSITOL 3-PHOSPHATE BINDING PROPERTIES OF PHAFIN2

Tuo-Xian Tang a, Mike Finlan a, and Daniel G. S. Capelluto a

Phafin2 is a phosphatidylinositol 3-phosphate (PtdIns(3)P)-binding protein involved in the regulation of endosomal cargo trafficking and the lysosomal induction of autophagy. Binding of Phafin2 to PtdIns(3)P is mediated by both its PH and FYVE domains. However, the structural bases for understanding how Phafin2 promotes signaling at endomembrane compartments are unknown. Here, we show that the isolated human recombinant Phafin2 is a moderately elongated monomer of ~28 kDa. Circular dichroism analysis indicates that Phafin2 exhibits an α/β structure and predicts ~40% of random coil content in the protein. Dynamic light scattering studies show that Phafin2 is relatively unstable, displaying a melting temperature of 35 °C. The folding-unfolding curves, obtained by the use of urea- and guanidine hydrochloride-mediated denaturation, indicate that Phafin2 undergoes a two-state native to denatured transition. Analysis of these transitions shows that the free energy change for urea- and guanidine hydrochloride-induced Phafin2 denaturation in water is ~4 kcal mol⁻¹. PtdIns(3)P binding to Phafin2 occurs with moderate affinity, triggering minor conformational changes in the protein. Taken together, these studies represent a platform for establishing the structural basis of Phafin2 molecular interactions and the role of the two potentially redundant PtdIns(3)P-binding domains of the protein in endomembrane compartments.

STAY CONNECTED:

bivt.edu

facebook.com/biocomplexity

twitter.com/biocomplexityvt

youtube.com/c/biocomplexity

google.com/+biocomplexity



Research Symposium | 2016-11-01 | Abstracts

A MATHEMATICAL-EXPERIMENTAL STRATEGY TO DECODE THE COMPLEX MOLECULAR BASIS FOR NEUTROPHIL MIGRATORY DECISION MAKING

Brittany Boribong¹, Sarah Kadelka², Mark Lenzi³, Stanca Ciupe, Ph.D.², Caroline N. Jones, Ph.D.³, and Liwu Li, Ph.D.³

¹Genetics, Bioinformatics, and Computational Biology, Virginia Tech

²Department of Mathematics, Virginia Tech

³Department of Biological Sciences, Virginia Tech

Chemoattractant signals are released from sites of infections and direct neutrophils to migrate toward the infection, where they initiate an immune response. Neutrophil chemotaxis can become disrupted during systemic inflammation occurring from sepsis. Consequently, neutrophils migrate toward healthy tissue and cause significant damage. To better study and understand migration patterns of neutrophils with single-cell resolution, we have developed a microfluidic device to measure neutrophil chemotaxis in an opposing chemoattractant gradient between fMLP (pro-resolution) and LTB4 (pro-inflammation) to model an immune response to sepsis. We also measured how neutrophil migration is affected when stimulated with pro- inflammation (LPS) and pro-resolution (LXA4) mediators. Our results have shown that a higher percentage of neutrophils migrate toward fMLP over LTB4, in both non-stimulated and stimulated neutrophils. LPS- stimulated neutrophils show a higher percentage of migration than non-stimulated neutrophils, whereas LXA4-stimulated neutrophils show a lower percentage. To understand why neutrophils favorably migrate toward fMLP over LTB4, we use a mathematical model to simulate the dynamics of neutrophil migration.

Using an ODE-based dynamical framework, we model the interaction of the mutually inhibitory GRK2 and GRK5 proteins and its role in neutrophil decision-making. Our model results show a bimodal switch between high and low levels of GRK2. We conclude from our experimental and computational model results that neutrophil migratory decision toward a specific chemoattractant in a competitive environment is based upon GRK2 expression levels. These studies could ultimately provide the basis for intervention strategies that would enable appropriate infiltration of phagocytes into inflammatory sites while minimizing neutrophil-mediated tissue injury.

**STAY
CONNECTED:**

bi.vt.edu

facebook.com/biocomplexity

twitter.com/biocomplexityvt

youtube.com/c/biocomplexity

google.com/+biocomplexity

Research Symposium | 2016-11-01 | Abstracts

COREG: IDENTIFICATION OF CO-REGULATORS IN GENOME SCALE TRANSCRIPTION REGULATORY NETWORKS

Qi Song^{1,2}, Song Li²

1. Ph.D. program in Genetics, Bioinformatics and Computational Biology, Virginia Polytechnic Institute and State University, Blacksburg, VA, 24061.
2. Department of Crop & Soil Environmental Sciences, Virginia Polytechnic Institute and State University, Blacksburg, VA, 24061. Email: alexsong@vt.edu; songli@vt.edu

Transcription factors tend to function as co-regulators to synergistically induce or inhibit expression of their targets in living organisms. In recent years, several genome-scale gene regulatory networks have been generated for multiple eukaryotic organisms including human, mouse, drosophila and Arabidopsis. However, existing module-finding algorithms fail to capture transcription co-regulators in these large-scale networks because these algorithms typically search for groups of densely connected genes (nodes) rather than co-regulating genes. In this study, we developed a new algorithm, CoReg, to identify transcription co-regulators in large-scale gene regulatory networks. CoReg groups genes based on weighted similarities between target nodes and regulatory nodes in a network. We applied hierarchical clustering followed by dynamic tree cut to identify co-regulatory modules. To compare the performance of our method with existing module-finding algorithms (Walk Trap, Edge Betweenness and Label Propagation), we conducted network-rewiring simulations and found that CoReg outperformed existing algorithms in identifying true co-regulators in large networks. We tested our algorithms in Arabidopsis Thaliana (*A. Thaliana*), Escherichia coli (*E. coli*) and Homo Sapiens (*H. Sapiens*) gene regulatory networks. We found that connectivity is high between network modules of co-regulators and their targets. In particular, we found that in the *A. Thaliana* regulatory network, the expression correlation for ~40% of the modules detected by CoReg are significantly higher than that of randomly constructed modules. This result strongly supports that the co-regulators identified by CoReg are involved in gene co-regulation. Our study provides a new tool for dissecting the architecture of regulatory network in multiple organisms.

STAY CONNECTED:

bivt.edu

facebook.com/biocomplexity

twitter.com/biocomplexityvt

youtube.com/c/biocomplexity

google.com/+biocomplexity



Research Symposium | 2016-11-01 | Abstracts

THE PROTECTIVE ROLE OF NLRX1 DURING INVASIVE PULMONARY ASPERGILLOSIS

Tariq Ayubi¹, Andrew Leber¹, Kelsey Simmons¹, Victoria Godfrey¹,
Raquel Hontecillas-Margazo¹, Josep Bassaganya-Riera¹, Shiv D. Kale¹
(1) Biocomplexity Institute of Virginia Tech.

Invasive pulmonary aspergillosis (IPA) is a high morbidity, high mortality fungal infection of the respiratory system effecting a diverse array of immunocompromised individuals. There is a pressing global need for the development of new antifungals due to growing incidences of resistant strains identified in clinical settings. The NLR (nucleotide-binding domain leucine-rich repeat containing) family of proteins are an essential component of the plant and animal immune response towards viruses, bacteria, and fungi. NLRX1 is a negative regulator of NF κ B signaling and plays an important role in human response to viruses and bacteria. Interestingly, Nlr1 was found to be upregulated in immune suppressed murine models of invasive pulmonary aspergillosis. Inoculation of Nlr1^{-/-} deficient mice with *Aspergillus fumigatus* resulted in significantly higher fungal loads in comparison to wildtype C57/BL6 mice in immunosuppressed models. A survival study indicated NLRX1 deficient mice were more susceptible to mortality earlier on during infection as well as an overall increase in mortality in an immunosuppressed mouse model. We utilized wildtype and nlr1^{-/-} bone marrow derived macrophages (BMDM) and bronchial airway epithelial cells (BEAS-2B) to further determine changes in conidial processing, and chemokine/cytokine signaling. Our findings suggest a minor increase in conidial processing, and elevated levels of known and novel immune signaling molecules during invasive pulmonary aspergillosis in the Nlr1^{-/-} BMDM. This study highlights the novel role of NLRX1 during IPA, identifies novel immune molecules that exacerbate inflammation during IPA, and a proof of concept for therapeutic development centered around NLRX1.

STAY CONNECTED:

bivt.edu

facebook.com/biocomplexity

twitter.com/biocomplexityvt

youtube.com/c/biocomplexity

google.com/+biocomplexity



Research Symposium | 2016-11-01 | Abstracts

DECISION-THEORETIC MODELS FOR HUMAN BEHAVIOR SIMULATION

Meghendra Singh, Achla Marathe, Samarth Swarup, Madhav Marathe
Network Dynamics and Simulation Science Laboratory

Large-scale simulations of human populations are a powerful tool for informed policy making. With respect to computational epidemiology such simulations are used to study disease outbreaks and assess intervention strategies. For such a simulation to be reliable, it is essential that the underlying models of agent behavior, capture the real-world decision making process of human beings. In this work we make use of the Semi-Markov Decision Processes (SMDPs) framework to construct decision-theoretic models of agent behavior grounded in data. We begin by identifying three categories of adaptive agent behaviors during disease outbreaks using results from an epidemiological survey. We incorporate these behaviors into the SMDP framework of states, options and rewards, contextualized in a compartmental epidemiology model. We optimize the option distributions on rewards to choose a candidate model that is consistent with the epidemiological survey data. A High Performance Computing oriented simulation is then used to implement the chosen model. The simulation can be used to study the impact of an epidemic on a very large human population.

STAY CONNECTED:

bivt.edu

facebook.com/biocomplexity

twitter.com/biocomplexityvt

youtube.com/c/biocomplexity

google.com/+biocomplexity



Research Symposium | 2016-11-01 | Abstracts

THE USE OF SOCIAL MEDIA BY THE U.S. ARMY

Ryan Whitcomb¹, Ian McCormack¹, Emma Baldwin¹, and Gizem Korkmaz²

¹Computational Modeling & Data Analytics, Virginia Tech

²Social and Decision Analytics Laboratory

Social media offers a new opportunity to capture soldiers' well-being and/or concerns about the Army. Data from various social media platforms can help to determine whether we can identify determinants of attrition. This study aims to understand the use of social media by the U.S. Army population. We ask whether there are differences in topics and sentiments across groups (e.g., soldiers vs. family) and across social media venues (e.g., Facebook, Twitter, YouTube), whether policy changes can be detected through the chatter in the social media and whether the sentiments towards these policies can be identified.

This study involves (i) identifying and developing a prioritization scheme for choosing social media sites relevant to the Army population to be used in the study, (ii) scraping the web to collect the relevant content, (iii) cleaning and preparing the data for analysis, (iv) conducting text mining and sentiment analysis relevant to the questions surrounding volunteer Army attrition. The findings will be presented at the poster session.

After further design, the project will proceed as an exploratory study. This is to account for the ambiguity of what data will be collected. Some ideas to be explored are post sentiment vs attrition and predicting how sentiments towards the Army will change according to the introduction of new policy. The exploratory process will show any interesting trends in the data that was scraped and form the direction that the concrete mathematical analysis will follow.

STAY CONNECTED:

bi.vt.edu

facebook.com/biocomplexity

twitter.com/biocomplexityvt

youtube.com/c/biocomplexity

google.com/+biocomplexity



Research Symposium | 2016-11-01 | Abstracts

CONSTRUCTING A BIOSYNTHETIC PATHWAY TO 3-HYDROXY-2-PYRROLIDINONE

Amanda Fisher

The specialty chemical 3-hydroxy-2-pyrrolidinone is the subject of biosynthetic pathway construction and metabolic engineering due to its favorable selling price (\$2,600/kg) and because there are no known routes for biological production. In order to produce 3-hydroxy-2-pyrrolidinone biologically, a new biosynthetic pathway was proposed, and enzyme promiscuity is now being engineered so that enzymes of the pathway can accept and convert non-native substrates. This is a significant challenge in biosynthetic pathway synthesis that will ultimately allow the production of new chemicals and provide more efficient routes to existing chemical targets. We are approaching the problem of biosynthetic pathway design using simultaneous computational and experimental approaches. In particular, molecular docking and dynamics simulations offer understanding of how native and non-native substrates interaction with active site residues of an enzyme. This approach is used to suggest enzyme engineering strategies that may be required for an enzyme to accept and convert a non-native substrate. In addition, a new enzyme activity screening technique has been developed using Raman spectroscopy that can probe the activity of engineered enzymes without enzyme purification or the development of specific colorimetric assays. Here, we also describe how these methods can be implemented in a semi-high-throughput assay to screen enzyme mutants for biosynthetic pathway construction.

STAY CONNECTED:

bivt.edu

facebook.com/biocomplexity

twitter.com/biocomplexityvt

youtube.com/c/biocomplexity

google.com/+biocomplexity



Research Symposium | 2016-11-01 | Abstracts

ATTRACTOR STABILITY IN FINITE ASYNCHRONOUS BIOLOGICAL SYSTEM MODELS

Ryan D. Pederson

In this paper we consider the following biological network models: the lac operon in *Escherichia coli* proposed in (Veliz-Cuba and Stigler, 2011) and the regulatory network involved in the control of the cell cycle and cell differentiation in the *Caenorhabditis elegans* vulva precursor cells proposed in (Weinstein et al., 2015). The authors in both cases evaluate their models using a parallel (synchronous) update schedule. We extend and complement these works by utilizing theoretical methods to efficiently identify all possible attractor configurations from these network models in the sequential update case. We find that the regulatory network model in lac operon can exhibit (under a select parameter configuration) four distinct possible limit cycle structures and the model in *C. elegans* can exhibit 125 distinct possible limit cycle structures under sequential updates. We analyze the variety and distribution of these limit cycle structures and comment on these results in the context of network model robustness. Furthermore, we demonstrate that the network model of *C. elegans* exhibits bistable long-term dynamics under select parameter and update schedule configurations. The techniques presented in this work are general and can be used to analyze other network models.

STAY CONNECTED:

bi.vt.edu

facebook.com/biocomplexity

twitter.com/biocomplexityvt

youtube.com/c/biocomplexity

google.com/+biocomplexity

Research Symposium | 2016-11-01 | Abstracts

FORECASTING THE RISK AND GEOSPATIAL DISTRIBUTION OF MELIOIDOSIS IN AUSTRALIA IN 2017

Pyrros A. Telonis

Melioidosis is a rare but exceedingly dangerous infectious disease found in Southeast Asia and Oceania. The etiological agent in question is the facultative intracellular bacterium *Burkholderia pseudomallei*. As saprophytes, the bacteria exist naturally in moist soils in the area. The majority of patients stricken with Melioidosis were exposed via inhalation or wound contamination, and almost all cases have occurred during the rainy season. The disease is also highly associated with extreme flooding and cyclones.

The pathogen itself is highly virulent, with an ID50 of just ten organisms when delivered via inhalation. Immunocompromising conditions, as well as alcoholism and uncontrolled diabetes, are significant risk factors for developing the disease, which often progresses to life threatening pneumonia. The incidence rate in Northern Australia hovers around 50 per 100,000 PY, while the case-fatality rate lies somewhere between 20-50% with intensive treatment.

The US Military has a surprising history with Melioidosis: the majority of recorded cases among servicemen were helicopter crews in Vietnam. It is thought that rotor wash agitated the moist soil which was then inhaled by the crew, leading to pneumonia. Though it is not a select agent, the Department of Defense (DOD) does consider the threat posed by Melioidosis when conducting exercises. Our study was prompted by a request from the DOD to determine which areas of Australia will be at high risk for Melioidosis in the coming rainy season (December 2016 to March 2017).

To predict areas of high risk, we used a modified ecological niche model. Typically niche modeling attempts to associate geographically heterogeneous independent variables, such as rainfall or temperature, with known presence and absence points. The output is a logistic function which can be used to predict presence or absence across the rest of the study area. We modified this procedure by calculating the incidence of specific regions during specific years, matching climatic data for those years, then using Boosted Regression Trees to model the relationship between the two. We then projected our model onto future climatic forecasts from the National Centers for Environmental Prediction. This yields a forecast of the severity and geographic spread of Melioidosis risk for the upcoming season.

STAY CONNECTED:

bivt.edu

facebook.com/biocomplexity

twitter.com/biocomplexityvt

youtube.com/c/biocomplexity

google.com/+biocomplexity



Research Symposium | 2016-11-01 | Abstracts

LINBASE - A WEB PORTAL FOR A NEW APPROACH TO BACTERIAL TAXONOMY BASED ON GENOME SIMILARITY

Long Tian¹, Grant Hughes², Lenwood S. Heath², Boris A. Vinatzer³

1 PhD program of Genetics, Bioinformatics and Computational Biology, Virginia Tech

2 Department of Computer Science, Virginia Tech

3 Department of Plant Pathology, Physiology and Weed Science, Virginia Tech

Bacterial taxonomy, i.e., the science of classifying, naming, and identifying bacteria, has been evolving with our understanding of life and along with new technology. Phytopathology, the study of plant diseases and the pathogenic agents that cause them, is an important field of research since it is concerned with crop protection and thus with feeding a growing world population. Phytobacteriologists are facing challenges when describing plant pathogens due to the controversial definition of “bacterial species”, which is the fundamental unit used to describe the diversity of life. Moreover, the current taxonomic system requires lengthy and laborious work to publish validly named species. Recently, genome sequencing has demonstrated its potential in efficiently describing microbial diversity and Average Nucleotide Identity (ANI) has become the new standard to determine how closely related bacteria are to each other. Life Identification Numbers (LINs), based on ANI, have proven to be a sound framework for identifying bacteria, for reflecting relationships between strains, and for giving stable genome similarity-based codes as names. Here we propose LINbase, a Web portal to a genomic similarity based approach to the taxonomy of bacteria. LINbase automatically classifies and assigns genome similarity-based codes, LINs, as names with a user-friendly interface at the front end, a fast and accurate calculation and processing pipeline at the back end, along with a fully structured and connected database for data extraction, transformation, and loading. In future versions, LINbase is going to provide the research community with an interactive environment for sharing and learning everything about bacteria.

STAY CONNECTED:

bivt.edu

facebook.com/biocomplexity

twitter.com/biocomplexityvt

youtube.com/c/biocomplexity

google.com/+biocomplexity



Research Symposium | 2016-11-01 | Abstracts

MODELING THE MECHANISMS BY WHICH HIV-ASSOCIATED IMMUNOSUPPRESSION INFLUENCES HPV PERSISTENCE AT THE ORAL MUCOSA

Meghna Verma¹, Samantha Erwin², Vida Abedi¹, Raquel Hontecillas¹, Stefan Hoops¹, Andrew Leber¹, Josep Bassaganya-Riera¹, and Stanca M Ciupe².

¹Nutritional Immunology and Molecular Medicine Laboratory, Biocomplexity Institute of Virginia Tech, Blacksburg, VA 24060

²Department of Mathematics, Virginia Tech, Blacksburg, VA 24060

These authors contributed equally to this work. Corresponding authors: Josep Bassaganya-Riera and Stanca M Ciupe. jbassaga@vt.edu (JBR), stanca@math.vt.edu (SC)

Human immunodeficiency virus (HIV)-infected patients are at an increased risk of co-infection with human papilloma virus (HPV), and subsequent malignancies such as oral cancer. To determine the role of HIV-associated immune suppression on HPV persistence and pathogenesis, and to investigate the mechanisms underlying the modulation of HPV infection and oral cancer by HIV we developed a mathematical model of HIV/HPV co-infection.

Our model captures known immunological and molecular features such as impaired HPV-specific effector T helper 1 (Th1) cell responses, and enhanced HPV infection due to HIV. We used the model to determine HPV prognosis in the presence of HIV infection, and identified conditions under which HIV infection alters HPV persistence in the oral mucosal system. The model predicts that conditions leading to HPV persistence during HIV/HPV co-infection are the permissive immune environment created by HIV and molecular interactions between the two viruses. The model also determines when HPV infection continues to persist in the short run in a co-infected patient undergoing antiretroviral therapy. Lastly, the model predicts that under efficacious antiretroviral treatment HPV infections will decrease in the long run due to the restoration of CD4+ T cell levels and protective immune responses.

Keywords: HIV, HPV, oral mucosa, mathematical models, combination antiretroviral therapy (cART), oral immune system, chronic HIV infection, oral co-infections

**STAY
CONNECTED:**

bivt.edu

facebook.com/biocomplexity

twitter.com/biocomplexityvt

youtube.com/c/biocomplexity

google.com/+biocomplexity



Research Symposium | 2016-11-01 | Abstracts

MATHEMATICAL MODELING OF CELLULAR PROCESSES

Jing Chen, Department of Biological Sciences

Our lab focuses on mathematical modeling of cellular processes. We are particularly interested in how temporal, spatial and mechanical regulations couple with biochemical signaling to realize cellular functions.

Currently, our lab works on two major projects. First, we are modeling circadian rhythmicity in mRNA polyA tail length. As a critical post-transcriptional modification, the polyA tail of mRNA controls mRNA stability and protein synthesis. Many mRNAs exhibit 24-hr rhythmicity in their polyA tail length. In addition, these mRNAs can be categorized into several general types according to the dynamic patterns of polyA tail length and mRNA level. Through modeling the polyA tail dynamics and protein dynamics, we aim to identify necessary conditions that give rise to each type of dynamic patterns, predict common regulatory steps shared among specific groups of mRNAs, and predict the relationship between polyA dynamics and protein dynamics.

Second, we are modeling coordinated motility in myxobacteria. To carry out complex “social” behaviors on the colony level, such as fruiting body formation, each myxobacterium glides on substrate with periodic reversals, and the reversal frequency is regulated by cell-cell contact. In each single cell, the periodic reversals are coupled with pole-to-pole oscillations of motility regulators. Lateral contacts are suggested to spatially align the motility regulators in contacting cells. We will develop an integrated model for myxobacteria motility that coherently links the motility regulation pathway to the spatiotemporal patterning of the motility regulators, and understand how intercellular contact confers intracellular signal to coordinate neighboring cells.

STAY CONNECTED:

bivt.edu

facebook.com/biocomplexity

twitter.com/biocomplexityvt

youtube.com/c/biocomplexity

google.com/+biocomplexity



Research Symposium | 2016-11-01 | Abstracts

HOW DO THE WNT FAMILY OF PROTEINS CONTRIBUTE TO ASTHMA AND ALLERGY?

Zoya Mahajan¹, Kelsey A. Simmons¹, TM Murali^{2,3}, Shiv D. Kale^{1,3}

- (1) Biocomplexity Institute of Virginia Tech
- (2) Department of Computer Science Virginia Tech
- (3) ICTAS Center for Systems Biology of Engineered Tissues

The Wnt family of secreted proteins, of which there are 19 in total for humans, are conserved across nearly all complex eukaryotic organisms, and are essential for organismal development and cellular homeostasis. Wnt proteins are known to signal through both the canonical β -catenin signaling path and other non-canonical paths. β -catenin signaling occurs in part through the TCF/LEF transcription factor, which plays an important role as a negative regulator of inter-cellular junction homeostasis. There is growing evidence suggesting that perturbation of endogenous Wnt signaling in the respiratory system of young children may correlate with the development of a number of diseases including allergy and asthma. Approximately 10.5 million children and 29.5 million adults are diagnosed with asthma and total medical expenditures exceed \$50 billion annually. We hypothesize that the mis-regulation of key Wnt proteins contributes to loss of cellular junctions between airway epithelial cells and results in increased fibronectin production by bronchial smooth muscle cells (BSMCs). Fibronectin is an emerging player of interest during allergy and asthma because an increase in the deposition of fibronectin into the sub-epithelial space of the airways is observed in all forms of asthma. We are utilizing PathLinker, an algorithm to compute the shortest paths within a network from a set of receptors to transcription factor receptors, to identify mechanisms of crosstalk between Wnt/ β -Catenin and the Glucocorticoid signaling pathways.

STAY CONNECTED:

bivt.edu

facebook.com/biocomplexity

twitter.com/biocomplexityvt

youtube.com/c/biocomplexity

google.com/+biocomplexity



Research Symposium | 2016-11-01 | Abstracts

CHARACTERIZING ADMINISTRATIVE DATA QUALITY - A NEW TOOL

Adrienne Rogers* (Virginia Tech) and Aaron Schroeder (Social and Decision Analytics Lab)

*Data Science for the Public Good Student Fellow, Social and Decision Analytics Lab

Nationally, states collect student records from all public schools for reporting and accountability purposes. These data are available through statewide longitudinal data systems (SLDS). States use SLDS data to conduct research to inform policy. The Social and Decision Analytics Laboratory, Biocomplexity Institute at Virginia Tech, developed a semi-automated R-markdown template to help researchers better understand the administrative data elements in which they are interested by using a disciplined process to identify data sources, assess the initial quality of these sources at the column/variable level, and understand what modifications may be necessary to prepare them for their intended research use(s). Working with the Virginia Department of Education (VDOE), VT iteratively tested and refined the template. The initial data framework summarizes a general approach for repurposing data, from discovery to analysis to inference. That framework was the culmination of a review and deep understanding of the data quality landscape, both across many fields and with respect to external data sources. Among the key challenges in using external data is the lack of control that the researcher has over the collection and processing of the data. The tool encompasses the creation of rules for each variable, the automation of the application of these rules across similar variables, and the implementation of an easy to use format for displaying the information. An example will be presented that identifies the rules for record consistency, including longitudinal consistency.

STAY CONNECTED:

bi.vt.edu

facebook.com/biocomplexity

twitter.com/biocomplexityvt

youtube.com/c/biocomplexity

google.com/+biocomplexity



Research Symposium | 2016-11-01 | Abstracts

SIMULATING WITHIN-VECTOR GENERATION OF THE MALARIA PARASITE DIVERSITY

Lauren M Childs, Department of Mathematics, Virginia Tech

Olivia F Prosper, Department of Mathematics, University of Kentucky

Plasmodium falciparum, the most virulent human malaria parasite, matures via asexual reproduction within the human host, but undergoes sexual reproduction within its vector host, the *Anopheles* mosquito. Consequently, the mosquito stage of the parasite life cycle provides an opportunity to create novel parasite genotypes in superinfected mosquitoes, increasing population diversity. Despite the important implications for disease transmission and malaria control, the mechanisms driving this diversity within the vector remain poorly understood. To examine the role that vector biology plays in modulating parasite diversity, we develop a two-part model that estimates the diversity as a consequence of different bottlenecks and expansion events occurring during the vector-stage of the parasite life cycle. For the underlying framework, we develop the first stochastic model of within-vector *P. falciparum* parasite dynamics and go on to simulate the dynamics of two competing parasite genotypes, emulating superinfection. We show that incorporating stochasticity is essential to capture the extensive variation in parasite dynamics, particularly in the presence of multiple genotypes. In contrast to predictions of deterministic models, we find that occasionally only genotypes with lower fitness survive to the sporozoite stage. This has important implications for onward transmission. The second part of our framework includes a model of sequence diversity generation resulting from recombination and reassortment between genotypes within a mosquito. Our model demonstrates that bottlenecks entering the oocyst stage decrease diversity from the initial gametocyte population in a mosquito's blood meal, but diversity increases with the possibility for recombination and proliferation in the formation of sporozoites.

STAY CONNECTED:

bivt.edu

facebook.com/biocomplexity

twitter.com/biocomplexityvt

youtube.com/c/biocomplexity

google.com/+biocomplexity

Research Symposium | 2016-11-01 | Abstracts

**FROM START TO FINISH: COMPUTATIONAL ANALYSIS OF CELL CYCLE CONTROL
IN BUDDING YEAST**

Pavel Kraikivski, Katherine C. Chen, Teeraphan Laomettachit, T.M. Murali and John J. Tyson

The cell division cycle is the ordered sequence of events by which a cell replicates its genome and segregates the replicated chromosomes to two daughter cells during mitosis. START and FINISH are the first and last events of this sequence, marking the cell's commitments, respectively, to begin a new round of DNA replication and to exit mitosis and divide. Based on the noteworthy progress made by molecular cell biologists in characterizing the genes and proteins that control cell cycle progression in budding yeast, we have built a comprehensive mathematical model of the molecular mechanisms underlying START and FINISH in yeast cells. For this detailed mathematical model, we use a new modeling framework in which all reactions are classified into three basic types: protein synthesis and degradation ($\square \rightarrow C \rightarrow \square$), phosphorylation and de-phosphorylation ($C \rightarrow CP$), and binding to activator or inhibitor partners ($C+A \rightarrow C:A$). The model is successful in explaining the observed phenotypes of a large number of START and FINISH mutants in budding yeast. We use the model to predict the phenotypes of novel combinations of mutant alleles. The credibility of these predictions has been assessed by robustness analysis, which distinguishes fragile predictions (which are sensitive to values of adjustable parameters) from robust predictions (properties that depend on the regulatory network itself rather than specific choices of parameter values). Future experimental tests of fragile predictions will be useful to constrain adjustable parameters of the model, and future tests of robust predictions will either confirm the underlying molecular mechanism or provide new insights into how the cell division cycle is regulated.

**STAY
CONNECTED:**

bi.vt.edu

facebook.com/biocomplexity

twitter.com/biocomplexityvt

youtube.com/c/biocomplexity

google.com/+biocomplexity



Research Symposium | 2016-11-01 | Abstracts

PREDICTING THE EFFECTS OF SENSE ANTISENSE TRANSCRIPTS INTERACTIONS IN A MATHEMATICAL MODEL OF CIRCADIAN RHYTHMS

Dorjsuren Battogtokh, Shihoko Kojima and John J. Tyson

Department of Biological Sciences, Virginia Polytechnic and State University, Blacksburg, VA, 24061

In our model of the mammalian circadian clock with Per2 and Per2AS interactions, we found a parameter set that fits well not only the experimental data reported in Ref [2] but also the timing patterns for expression of core clock genes in experimental data and recent model [1]. To explore the model further and make testable predictions, we computed detailed one- and two-parameter bifurcation diagrams for the model [3]. Using these theoretical analysis, we plan to test the model's predictions experimentally and decipher the molecular mechanisms by which Per2AS regulates the circadian clock.

1. A. Relogio et al. (2011), *Tuning the Mammalian Circadian Clock: Robust Synergy of Two Loops*, *PLoS One*, 7(12):e1002309

2. Koike, N. et al. *Transcriptional Architecture and Chromatin Landscape of the Core Circadian Clock in Mammals*. *Science*, doi:10.1126/science.1226339 (2012).

3. D. Battogtokh et al. *Bifurcation analysis and simulations of a mathematical model of mammalian circadian rhythm*, preprint 2016.

**STAY
CONNECTED:**

bivt.edu

facebook.com/biocomplexity

twitter.com/biocomplexityvt

youtube.com/c/biocomplexity

google.com/+biocomplexity



Research Symposium | 2016-11-01 | Abstracts

UNDERSTANDING HEALTH DISPARITIES IN AN INFLUENZA EPIDEMIC

Lijing Wang, Jiangzhuo Chen, Achla Marathe

Network Dynamics and Simulation Science Laboratory, Biocomplexity Institute, Virginia Tech, Blacksburg, VA

Infectious disease epidemics such as influenza and Ebola pose a serious threat to global public health. However certain demographics and cohorts are at a higher risk of infection than others due to a variety of factors. In this research we explore the heterogeneities in social and personal attributes as the cause of health disparities, in an influenza epidemic. We use an agent-based model to simulate an influenza epidemic over a synthetic social contact network of Montgomery county population in Southwest Virginia. We divide the population into age and income based cohorts and measure the direct and indirect economic impact from vaccination based interventions for each cohort. The metrics used for measuring the economic impact are net return per capita, net return per vaccinated person and net return per dollar spent. Results show significant disparities in health outcomes and in economic outcomes across age groups and income groups. We designed several public health vaccination intervention strategies, and have identified the optimal among them. The findings have significant policy implications that may assist public health decision making in assigning limited intervention resources.

STAY CONNECTED:

bi.vt.edu

facebook.com/biocomplexity

twitter.com/biocomplexityvt

youtube.com/c/biocomplexity

google.com/+biocomplexity



Research Symposium | 2016-11-01 | Abstracts

TITLE: EGR1 RECRUITS TET1 TO SHAPE THE BRAIN METHYLOME DURING EARLY POSTNATAL DEVELOPMENT

Zhixiong Sun^{1,2}, Ming-an Sun¹, Xiguang Xu^{1,2}, Jiyoung Lee^{1,8}, Xia Wang³, Keqiang Cheng², Jiang Xi⁴, Michelle Theus³, Liwu Li², Jinsong Zhu⁹, Alexei Morozov^{5,6,7}, Jianjun Chen⁴, Hehuang Xie^{1,2}

1Epigenomics and Computational Biology Lab, Biocomplexity Institute of Virginia Tech

2Department of Biological Sciences

3Department of Biomedical Sciences & Pathobiology, Virginia Tech

4Department of Cancer Biology, University of Cincinnati, Cincinnati, OH

5Virginia Tech Carilion Research Institute

6School of Biomedical Engineering and Sciences

7Department of Psychiatry and Behavioral Medicine, Virginia Tech Carilion School of Medicine, Roanoke, VA

8GBCB program, Virginia Tech

9Department of Biochemistry, Virginia Tech

Early growth response gene-1 (Egr1) is a critical transcription factor involved in many important biological processes, including neuronal plasticity and memory formation. With a rapid increase in expression during the first few weeks after birth, Egr1 controls the selection, maturation and functional integration of newborn neurons. The regulation of Egr1-mediated gene expression has been shown to be under methylation control. However, Egr1 target sites and their epigenetic regulation in the nervous system remains largely unknown. Recently, we observed a large number of genomic loci with their cell-type specific methylation patterns established during postnatal frontal cortex development. For both human and mouse, these loci enrich for transcription factor binding motifs, in particular for Egr1. We have now performed ChIP-seq to determine EGR1 binding sites in mouse frontal cortex and confirmed that many EGR1 binding sites become hypo-methylated in mature neurons but remain heavily methylated in glia. In addition, our preliminary data suggested that EGR1 physically interacts with and recruits DNA demethylation enzyme, TET1, to its binding sites for demethylation and activation of EGR1 downstream genes. This study provides important new insights into how early life experience shapes the brain methylome and reveals a key epigenetic mechanism that underlie postnatal brain development.

**STAY
CONNECTED:**

bivt.edu

facebook.com/biocomplexity

twitter.com/biocomplexityvt

youtube.com/c/biocomplexity

google.com/+biocomplexity

Research Symposium | 2016-11-01 | Abstracts

EFFECT OF MODELING SLUM POPULATIONS ON INFLUENZA SPREAD AND CONTROL IN DELHI

Presenting Author: Shuyu Chu

Co-authors: A. Adiga, J. Chen, S. Chu, Y. Chungbaek, S. Eubank, S. Gupta, M. Khan, C. Kuhlman, B. Lewis, A. Marathe, M. Marathe, H. Mortveit, S. Swarup, A. Vullikanti, A. Wilson, and D. Xie

Network Dynamics and Simulation Science Laboratory, Biocomplexity Institute, Virginia Tech, Blacksburg, VA

Objectives This research studies the impact of Influenza epidemic on the slum and non-slum populations of Delhi, India, by taking proper account of slum demographics and residents' activities, using a highly resolved social contact network of the 13.8 million residents of Delhi.

Methods An SEIR model is used to simulate the spread of Influenza on two different synthetic social contact networks of Delhi, one where slums and non-slums are treated the same in terms of their demographics and daily sets of activities and the other, where slum and non-slum regions have different attributes.

Results Differences between the epidemic outcomes on the two networks are large. Time-to-peak infection is overestimated by several weeks, and the cumulative infection rate and peak infection rate are underestimated when slum attributes are ignored.

Conclusions Slum populations play a significant role on the spread and control of influenza in urban areas. Improper specification of slums in large urban regions results in underestimation of infections in the entire population. The most effective intervention method is to give vaccines to slum dwellers and apply social distancing to non-slum people.

**STAY
CONNECTED:**

biv.tedu

facebook.com/biocomplexity

twitter.com/biocomplexityvt

youtube.com/c/biocomplexity

google.com/+biocomplexity



Research Symposium | 2016-11-01 | Abstracts

UNDERSTANDING SOCIAL DIFFUSION DYNAMICS USING NETWORKED COGNITIVE SYSTEMS

Daniel Chen

Binary decision models with externalities is a class of models that describe individuals making a simple binary choice, whose decision depends on other people. This class of models is generalizable to infectious diseases, decision making, information spread, fads, riots, etc. However, binary decision models are not rich enough to describe the complex process of human attitude formation that eventually lead to behaviors.

The Theory of Reasoned Action is a health behavior model that states behaviors stem from a series of beliefs. This theory is used as the computational framework for constraint-satisfaction artificial neural-networks which allow us to model psychologically plausible decisions on an individual level. This model can be combined with agent-based models, which model health behavior at a population level in order to add psychology into a dynamic social network framework. This new framework for epidemiological modeling of health behaviors where the agents may exhibit psychological plausible decisions allows us to study the spread of ideas within a social network. This will give us insights as to what drives behavioral risk factors for population health behaviors.

STAY CONNECTED:

bivt.edu

facebook.com/biocomplexity

twitter.com/biocomplexityvt

youtube.com/c/biocomplexity

google.com/+biocomplexity



Research Symposium | 2016-11-01 | Abstracts

A MODEL OF GRANULOCYTE MONOCYTE PROGENITOR DIFFERENTIATION

Bronson Weston¹, Liwu Li², and John J. Tyson²

¹Genetics, Bioinformatics, and Computational Biology, Virginia Polytechnic Institute and State University

²Department of Biological Sciences, Virginia Polytechnic Institute and State University

Granulocyte-monocyte progenitor (GMP) cells typically differentiate into monocytes or granulocytes depending on cytokine exposure and their relative concentrations. A third cell type within this myeloid lineage has been observed which have been named myeloid derived suppressor cells (MDSC). MDSCs make up a heterozygous population of naive cells that have characteristics of both granulocytes and monocytes and have been associated with suppression of T cell immune response and aiding tumor growth. We have constructed a novel motif composed of critical transcription factors and receptors in the differentiation process of GMP cells. Through dynamic systems theory we have modeled this differentiation process in response to cytokines such as GM-CSF, M-CSF, and G-CSF. We show how GM-CSF can induce populations to favor monocyte or granulocytes depending on concentration, and offer an intracellular explanation for this intriguing behavior. We demonstrate that GM-CSF can be paired with G-CSF to reduce the concentration of G-CSF required to induce differentiation of granulocytes and we consider the role GM-CSF could play in treatment of neutropenia. Finally, we show that under unique conditions it may be possible for GMP cells to differentiate into a hybrid phenotype which is representative of MDSCs.

STAY CONNECTED:

bi.vt.edu

facebook.com/biocomplexity

twitter.com/biocomplexityvt

youtube.com/c/biocomplexity

google.com/+biocomplexity



Research Symposium | 2016-11-01 | Abstracts

SIMULATION SUMMARIZATION

Nidhi Parikh, Madhav Marathe, and Samarth Swarup

Large-scale and complex multiagent simulations are becoming common. They can generate much more data in each simulation run than goes into the simulation. Hence, new methods are becoming necessary to make sense of the results of these simulations. Even concisely summarizing the results of a given simulation run is a challenge. Here we present the problem of simulation summarization: how to extract the causally-relevant descriptions of the trajectories of the agents in the simulation. We present a simple algorithm to compress agent trajectories through state space by identifying the state transitions which are relevant to determining the distribution of outcomes at the end of the simulation. We apply it to a large-scale simulation of the aftermath of a disaster in a major urban area and show that it can identify a number of meaningful states, for example, it identifies being in a healthcare location or out of the area as states that improve health outcomes while panic, household reconstitution, and healthcare-seeking as states (behaviors) that worsen health.

STAY CONNECTED:

bivt.edu

facebook.com/biocomplexity

twitter.com/biocomplexityvt

youtube.com/c/biocomplexity

google.com/+biocomplexity



Research Symposium | 2016-11-01 | Abstracts

MODELING THE NUTRIENT SIGNALING NETWORK IN SACCHAROMYCES CEREVISIAE

Amogh Jalihal¹, Pavel Kraikivski, Ph.D.², T.M. Murali, Ph.D.³, and John Tyson, Ph.D.²

¹Genetics, Bioinformatics and Computational Biology, Virginia Tech

²Department of Biological Sciences, Virginia Tech

³Department of Computer Science, Virginia Tech

The regulation of cell growth and division has long been considered a central problem in cell biology. Much work has been done both theoretically and experimentally towards expanding our understanding of the processes of cell division and cell size sensing. However, there has been almost no systematic investigation of the networks regulating cellular growth rates. Due to the complex nature of the molecular regulatory mechanisms governing these processes, mathematical modeling has become a powerful tool that has been used to study the dynamic responses of these regulatory networks to intra- and extracellular perturbations. Here, we propose an ODE-based dynamical model of the regulatory mechanism governing growth in yeast that will simulate variations in cellular growth rates as a function of varying macronutrient inputs, namely carbon and nitrogen sources. This network will be composed of components from the nutrient sensing and signaling pathways in yeast that ultimately modulate overall protein synthesis in the cell, in turn affecting the mass growth rate. Further, we plan on integrating this model of nutrient signaling with the highly successful model of the yeast cell cycle, with the goal of refining the phenotypic predictions made by the cell cycle model. By incorporating into this model the diverse nutritional contexts in which experiments are carried out, we aim to predict a wider range of nutrient-sensitive phenotypes ranging from the diauxic shift to nutrient starvation.

STAY CONNECTED:

bivt.edu

facebook.com/biocomplexity

twitter.com/biocomplexityvt

youtube.com/c/biocomplexity

google.com/+biocomplexity



Research Symposium | 2016-11-01 | Abstracts

A SEMANTIC NETWORK APPROACH TO UNDERSTANDING VACCINE HESITANCY ON THE WEB

Gloria Kang^{1, 2*}, Sinclair Ewing-Nelson², Lauren Mackey², Achla Marathe², Samarth Swarup²

1Department of Population Health Sciences, Virginia-Maryland College of Veterinary Medicine; 2Network Dynamics and Simulation Science Laboratory

*Presenter

Vaccine hesitancy has existed for as long as vaccines have existed - with individual reasons for accepting, refusing, or delaying vaccination changing over time. While the advent of online social media has provided new platforms for sharing information – it has simultaneously created challenging new backdrops for emerging issues of public health and communication. Our study objectives were to (1) identify popular sources of vaccine information circulating on Twitter; (2) create graph-based representations of vaccine beliefs and sentiment extracted from (1); and (3) analyze the structural properties of these networks differing in vaccine sentiment through semantic network analysis. The resulting network topology and its measures helped elucidate the most salient concepts within each network of vaccine beliefs. Networks representing positive, negative, and neutral sentiment toward vaccines exhibited stark contrast involving semantic differences (something which may be lost when relying on traditional natural language processing methods). Overall, this study demonstrated the unique potential for semantic network analysis, as part of a much-needed interdisciplinary effort, to enhance meaningful understanding of highly-complex public health issues such as vaccine hesitancy.

**STAY
CONNECTED:**

bivt.edu

facebook.com/biocomplexity

twitter.com/biocomplexityvt

youtube.com/c/biocomplexity

google.com/+biocomplexity



Research Symposium | 2016-11-01 | Abstracts

NETWORK SECURITY DATA ANALYTICS

Mark E. DeYoung, Information Technology Security Laboratory (ITSL) Virginia Polytechnic Institute and State University, Blacksburg, VA; Air Force Institute of Technology (AFIT), Wright-Patterson AFB, Dayton OH

Data-driven network security and information security efforts have decades long history. A deluge of logged events from network mid-points and end-points, coupled with unprecedented temporal depth in data retention, are driving an emerging market for automated cognitive security products. Historically, new technologies like this are delivered as non-contextualized black boxes. We introduce network security data analytics methodology within the context of intelligence activities and products. Our initial efforts focused on collection, aggregation, and normalization of data sets. Our central investigative question is: what can we learn from retrospective analysis of predictive factors (within probabilistic bounds) from historical logged events? We are now moving toward model development to discover single-event instance level characteristics (syntactic characteristics) and multi-event/inter-instance structural relationships (i.e. semantic, temporal, spatial, and other unobserved relationships). Our intuition is that much of the logged event data will involve workloads that are generally arbitrarily separable with short range (temporal and spatial) inter-event data dependencies. These workload structures should be amenable to parallel algorithms. We use a flexible processing architecture with substitutable and composable components to ingest, process, and analyze logged events. We are evaluating various design variations (substituting components) and design iterations (tuning component parameters). We propose network security data analytics as a framework to develop and evaluate cognitive security products that can satisfy operational needs.

STAY CONNECTED:

bi.vt.edu

facebook.com/biocomplexity

twitter.com/biocomplexityvt

youtube.com/c/biocomplexity

google.com/+biocomplexity



Research Symposium | 2016-11-01 | Abstracts

RESIDENTIAL SMOKE ALARM NEED IN ARLINGTON COUNTY

Mark Almanza* (Virginia Tech), Will Sandholtz* (UC Berkeley), Sam Tyner* (Iowa State), Gizem Korkmaz, Joshua Goldstein, and Vicki Lancaster (Social and Decision Analytics Lab)

*Data Science for the Public Good Student Fellows, Social and Decision Analytics Lab

In 2014, U.S. fire departments responded to an estimated 367,500 residential fires that resulted in 3,275 civilian deaths and 15,775 injuries and an estimated \$6.8 billion in property loss in home fires. 1 Home fires caused 2,745, or 84%, of the civilian fire deaths. The use of smoke detectors has been shown to be an effective, reliable, and inexpensive method of providing early warning in residential fires. 2 It is found that 70% of home fire deaths occur in homes with either no smoke alarm or homes in which none of the smoke alarms sounded, and that working smoke alarms can decrease the risk of death by 40%-50%.³

Arlington County Fire Department (ACFD) initiated the Operation FireSafe program in June 2015 to increase the number of residences in the county with smoke alarms. ACFD visited 5,623 single family homes countywide to offer free smoke alarm inspections and installations. Building a model to identify single family homes that are in need of an alarm may increase the efficiency of the program.

In this study, we combine the information collected by ACFD through the Operation FireSafe program and the housing characteristics of single family homes in Arlington County obtained from the CoreLogic dataset to develop a model (i) to identify home characteristics that can be used as predictors of smoke alarm absence and (ii) to compute the likelihood of having a smoke alarm, for all single family homes in Arlington County to inform future Operation FireSafe visits. We find that the value and the age of the single family home are significant predictors of smoke alarm presence/need. We also calculate the predicted probabilities of smoke alarm presence for 49,178 residential units in Arlington County. We observe that the single family homes in south-eastern regions of Arlington are more likely to be in need of smoke alarms.

1 Haynes, H. J. (2015). *Fire loss in the United States during 2014*. National Fire Protection Association. *Fire Analysis and Research Division*.

2 Marshall, S. W., Runyan, C. W., Bangdiwala, S. I., Linzer, M. A., Sacks, J. J., & Butts, J. D. (1998). *Fatal residential fires: Who dies and who survives?* *JAMA*, 279(20), 1633-1637.

3 Ahrens, M. (2004). *US experience with smoke alarms and other fire detection/alarm equipment*. NFPA, Quincy, MA.

STAY CONNECTED:

bivt.edu

facebook.com/biocomplexity

twitter.com/biocomplexityvt

youtube.com/c/biocomplexity

google.com/+biocomplexity



Research Symposium | 2016-11-01 | Abstracts

EVALUATING INNOVATION OF OPEN SOURCE SOFTWARE

John Higgins¹, Alex Gagliano¹, Romcholo Macatula¹, Gizem Korkmaz² and Stephanie Shipp²

¹Computational Modeling & Data Analytics, Virginia Tech

²Social and Decision Analytics Laboratory

The National Science Foundation is interested in measuring innovation using non-survey sources of data that will allow them the opportunity to explore innovation in new areas, such as the contribution due to Open Source Software. This research aims to develop a series of metrics to quantify this contribution. Methods include regression, Markov chains, and several unsuper-vised methods to study the spread of and current trends in Open Source Softwares. Data will be scraped from various sites that compile information about Open Source Softwares. Overall, this research will be a first step for the National Science Foundation's adaptation into funding Open Source Softwares.

STAY CONNECTED:

bi.vt.edu

facebook.com/biocomplexity

twitter.com/biocomplexityvt

youtube.com/c/biocomplexity

google.com/+biocomplexity

Research Symposium | 2016-11-01 | Abstracts

COMPARATIVE GENOMICS INSIGHTS INTO SPECIATION AND EVOLUTION IN DROSOPHILA

Lin Kang

Speciation and adaptation has always been of great interest to biologist. The Hawaiian archipelago provides a natural arena for understanding adaptive radiation and speciation, and genomics and bioinformatics offer new approaches for studying of this. A pervasive influence of positive selection on divergence of three species Hawaiian pictured-winged *Drosophila* species, *Drosophila planitibia*, *Drosophila heteroneura*, and *Drosophila silvestris* was found with the signatures of positive selection more prominent in sympatry than allopatry. Sequence variation in Olfactory receptor and Gustatory receptor genes seems to play a major role in adaptive radiation in Hawaiian *Drosophila*. Further analysis of RNAseq data from *Drosophila planitibia*, *Drosophila silvestris*, and their sterile F1 hybrid (three distinct sperm phenotype: motile sperm, no sperm, and immotile sperm) shows many genes significantly differentially expressed between backcross males with no sperm compared with those backcross males with motile sperm and immotile sperm, and all of those genes were underexpressed in males with no sperm, including a number of genes with previously known activities in adult testis. An allele-specific expression analysis showed overwhelmingly more cis-divergent than transdivergent genes, with no significant difference in the ratio of cis- and trans-divergent genes among the sperm phenotypes. Additional experimental evolution study analyzing genomes of *Drosophila melanogaster* from lines evolving under long-term directional selection for increased desiccation resistance in comparison with control lines illustrates extensive footprints on the genome with a high degree of fixation of alleles in surrounding neighborhoods of eroded heterozygosity, as well as the high divergence within coding sequences between selected and control lines. It indicates that big impact on genome diversity many organisms may experience to adapt to new environmental setting like dry environment.

**STAY
CONNECTED:**

bivt.edu

facebook.com/biocomplexity

twitter.com/biocomplexityvt

youtube.com/c/biocomplexity

google.com/+biocomplexity



Research Symposium | 2016-11-01 | Abstracts

THE PREDICTED EFFECTS OF TOPOISOMERASE II ON 3D NUCLEIC ARCHITECTURE

Kyle Titus-Glover

A cell's ability to pack an immense length of DNA into its nucleus' small volume is a relatively poorly understood phenomenon that continues to baffle scientists and researchers. The mysterious organization of this "packing" is crucial to many cellular processes as the access to and expression of certain genes are necessary for the cell's function and placement in the body. Therefore, the analysis of this singularity, being the complexity of the folds of the chromosomes, becomes key in order for scientists to truly understand cellular processes in general. This project investigates the effects of the Topoisomerase II enzyme on the 3D architecture of the fruit fly genome. Topoisomerases are enzymes in the cell that regulate the abundant number of chromosomal coils that form in the nucleus. In this project, we investigate the changes to 3D paradigms constructed in the image of a fruit fly's genome and, depending on the level of topoisomerase, determine whether the chromosomes are bound to the nuclear surface. The project utilizes molecular dynamics simulations of coarse-grained models of chromosomes in fruit flies through the use of well-established codes in Espresso and Matrix Laboratory (MatLab). The outcome of these results indicate a faster deterioration of distinct chromosomal territories in the presence of the enzyme and propose a possible correlation between topoisomerase activity and cellular aging. Such findings suggest an immense significance of the enzyme as not only the architect of the aforementioned enigmatic designs of the chromosomes but also as a regulator for stages within the cell's lifespan.

STAY CONNECTED:

bi.vt.edu

facebook.com/biocomplexity

twitter.com/biocomplexityvt

youtube.com/c/biocomplexity

google.com/+biocomplexity



Research Symposium | 2016-11-01 | Abstracts

EPIGRIND: ADAPTIVE MODELS OF ZIKA TRANSMISSION

James Schlitt

Following the distressing trend set by its viral cousins, Zika virus (ZIKV) has crossed the Pacific and emerged as a threat in the Americas. Though illness caused by ZIKV is thought to be milder than those caused by its relatives such as dengue and yellow fever, it is far from harmless. Implicated in severe intrauterine infections, ZIKV has been identified as a cause for the increase in reports of extremely debilitating microcephaly in Brazil.

In this study we fit a metapopulation patch model to the epidemic curves of ZIKV infections in the 27 administrative regions of Brazil. We used said regions as our metapopulation patches, and within each fit a compartmental SEIR model for both humans and mosquitoes based on the work of Manore et al. (2015). Mosquito density was estimated using ecological niche models of the appropriate *Aedes* species from Walter-Reed Biosystematics Unit, while human mobility between patches was estimated with a gravity model of transportation.

The resulting model shows the predicted spread of ZIKV across Brazil with an estimated 371,660 cases by June 3rd of 2016. To address the threat of ZIKV in the United States, we are using a model calibrated to the Wynwood, Miami cluster of autochthonous ZIKV transmission to model scenarios within the Tidewater, Virginia region via simulated local travel patterns.

STAY CONNECTED:

bivt.edu

facebook.com/biocomplexity

twitter.com/biocomplexityvt

youtube.com/c/biocomplexity

google.com/+biocomplexity

Research Symposium | 2016-11-01 | Abstracts

DEALING WITH INSTABILITY: HOW CYTOKINESIS FAILURE CAN CONTRIBUTE TO GENOME EVOLUTION AND DIVERSITY IN COLORECTAL CANCER CELLS

Nicolaas Baudoin

Tetraploid cells (those with double the normal genome content) are found in a variety of animal tissues under normal and malignant conditions. These cells have been shown to be more tumorigenic in mice compared to isogenic diploid counterparts, and are thought to play significant role in tumorigenesis. The specific role that tetraploidy plays in cellular evolution, however, remains unresolved. Two leading ideas for tetraploidy's role in allowing rapid evolution are that (1) the processes that lead to tetraploid cells give rise also to supernumerary centrosomes (causing perpetual genomic instability by altering the mitotic spindle and chromosome segregation dynamics), and (2) that the extra chromosomes allow cells to tolerate greater genomic re-arrangements to find new, fitter combinations in a changing environment. Here, we examine the early evolution of karyotype and centrosome number in newly formed tetraploid cells. Multipolar cell division is common in early divisions, and these events lead to asymmetric distribution of the genome to two or more daughter cells. We find that the number of cells with supernumerary centrosomes (and the rates of multipolar divisions) rapidly decreases during the first week of post-tetraploid growth, and return to levels comparable to the diploid within 12 days of evolution. Karyotype analysis indicates that intermediate (sub-tetraploid, but not normal diploid) karyotypes generated from these divisions are also rapidly eliminated from the proliferating cell population. Live cell imaging of cells containing fluorescently marked centrosomes indicate that asymmetric clustering of centrosomes during bipolar mitoses could generate a population of near-tetraploid cells that contain a normal number of centrioles in G1. This near-tetraploid population of cells shows an increased ability to grow in an anchorage-independent manner. We conclude that the pressures of mitotic stability and genome integrity limit the success of cells with supernumerary centrosomes and/or highly abnormal karyotypes, and these traits are therefore quickly selected against. Furthermore, our results show that this evolved population of post-tetraploid cells has a survival advantage compared to diploid cells under certain stress conditions.

**STAY
CONNECTED:**

bivt.edu

facebook.com/biocomplexity

twitter.com/biocomplexityvt

youtube.com/c/biocomplexity

google.com/+biocomplexity



Research Symposium | 2016-11-01 | Abstracts

SEASONAL EPIDEMIC FORECASTING BY SMART BEAM-PARTICLE FILTER FRAMEWORK

Farzaneh Tabataba

Numerous computational and mathematical methodologies have been proposed for forecasting seasonal epidemics over the past decades. Huge pandemics like 2009 H1N1, Ebola, and Zika in recent years made the scientist more ambitious to improve the previous methods and alleviate their problems to provide well-informed and reliable forecasts. The output of the forecasting models is of prime importance to epidemiologists and health-care providers.

In this presentation, I will describe a smart Beam-Particle Filter calibration framework to find the best configuration of a data-driven model for the diseases like flu and Ebola. We model the infectious disease dynamics by an Agent-Based framework which simulated the spread of disease over a synthetic contact network. We will show the robustness of the methodology by experimental test on synthetic data generated for influenza. We also evaluate the performance of the forecast method by applying it to the synthetic Ebola disease dataset that is provided by National Institutes of Health as the real data for Ebola forecasting challenge.

STAY CONNECTED:

bi.vt.edu

facebook.com/biocomplexity

twitter.com/biocomplexityvt

youtube.com/c/biocomplexity

google.com/+biocomplexity



Research Symposium | 2016-11-01 | Abstracts

COMPUTATIONAL CONSTRUCTION OF TOXICANT RESPONSE NETWORKS

Presenting author: Jeffrey Law - Genetics, Bioinformatics and Computational Biology, VT

Other authors: Sophia Orbach - Chemical Engineering, VT; Bronson Weston - Genetics, Bioinformatics and Computational Biology, VT; T.M. Murali - Department of Computer Science, VT

Efforts such as the EPA's ToxCast seek to quickly and efficiently screen thousands of chemicals for potential human and environmental effects. For each chemical, these efforts study a fixed set of proteins and pathways and they do not consider the complex network of interactions among assayed proteins and pathways. We seek to connect the responding proteins in the context of the underlying network of regulatory and physical interactions.

Our approach is to compute the k shortest paths (using PathLinker) from perturbed receptors to perturbed transcription factors (TFs) in the human interactome to build a response network for each toxicant. These toxicant response networks may reveal important intermediate proteins that have not have been tested and physiological processes that have not been previously implicated in connection with the chemical.

We specifically analyze the response network of Lovastatin, a drug used to lower cholesterol. We compute functions enriched in the set of recovered proteins returned by PathLinker, visualize the response network in GraphSpace, and show literature support for our predictions. PathLinker identified pathways known to be perturbed by Lovastatin, as documented in the literature. Enriched pathways with no literature support may represent previously unknown effects of Lovastatin. These results suggest that PathLinker may be a promising method for constructing response networks for less well studied toxicants.

Our ongoing research is to apply PathLinker to other chemicals and show statistical significance for toxicant response networks.

STAY CONNECTED:

bi.vt.edu

facebook.com/biocomplexity

twitter.com/biocomplexityvt

youtube.com/c/biocomplexity

google.com/+biocomplexity



Research Symposium | 2016-11-01 | Abstracts

EFFECTS OF NETWORK STRUCTURE ON EPIDEMIC MODELLING

Madhurima Nath, Yihui Ren, Yasamin Khorramzadeh and Stephen Eubank

The effect of the topology of a finite sized interacting system, modelled into a network, on its dynamics is an interesting question. Various methods have been proposed in the literature for building different equivalence classes or families of networks based on the structural aspects of the system. It is observed that similarities in the local statistics of two networks are not sufficient to predict the dynamics on them. We suggest a global statistic, the Moore and Shannon's network reliability polynomial that depends on both the structure and the dynamics to explore the behaviour of a diffusive process on a network. It gives the probability that a system composed of many different interacting components has a desired property. The computation of the reliability polynomial exactly is often NP-hard, but estimating it using Monte-Carlo simulation is feasible even for graphs with hundreds of millions of edges. The probabilistic nature of the polynomial allows the mapping of parameters of one network on to another using a simple transformation that keeps the dynamical phenomenon invariant.

STAY CONNECTED:

bivt.edu

facebook.com/biocomplexity

twitter.com/biocomplexityvt

youtube.com/c/biocomplexity

google.com/+biocomplexity



Research Symposium | 2016-11-01 | Abstracts

SINGLE-CELL ANALYSIS REVEALS THAT SILVER NANOPARTICLE EXPOSURE LEADS TO MULTI-NUCLEATION THROUGH DEFECTIVE CELL DIVISION

Ellen Garcia¹, Cindy Alms², Conor Kelly¹, Adam Smith¹, Linsey Marr³, and Daniela Cimini¹

¹Department of Biological Sciences and Biocomplexity Institute, Virginia Tech, Blacksburg, VA 24061; ²University of Maryland, Baltimore County; ³Department of Civil and Environmental Engineering, Virginia Tech, Blacksburg, VA

Multiple agencies, including the U.S. Environmental Protection Agency and the National Academy of Science, are urging for a radical paradigm shift from standard, whole animal testing to alternative and novel technologies. To meet this urgent need, we aimed to develop a new, cell division-focused toxicity assay by investigating the mechanism of toxicity from silver nanoparticles (AgNPs) on human retinal pigment epithelial (RPE-1) cells. Cultured RPE-1 cells were treated with varying concentrations of AgNPs and live-cell microscopy was used to analyze the behavior of cells undergoing cell division over a 24 hour time period. A physical interaction between cells and particles was visually observed and 100% of treated cells appeared to engulf particles. We found that higher concentrations of AgNPs resulted in large numbers of cells stalling in mitosis and/or dying. In contrast, untreated cells displayed normal mitotic behavior. High-resolution fluorescence microscopy performed in chronically treated cell populations identified an increased percentage of binucleated cells. Further live-cell analysis indicated that one major cell division defect could explain the binucleated cell phenotype. Indeed, treated cells failed cytokinesis (cytoplasmic division following mitotic chromosome segregation) more often than control cells. Overall, our results indicate that AgNPs specifically impair cell division, not only further confirming toxicity to human cells, but also revealing specific, previously unreported toxicity mechanisms and highlighting the propagation of adverse phenotypes within the cell population after exposure. Furthermore, this work illustrates that cell division-based assays and single-cell analysis could greatly benefit chemical safety experimentation in the future.

**STAY
CONNECTED:**

bivt.edu

facebook.com/biocomplexity

twitter.com/biocomplexityvt

youtube.com/c/biocomplexity

google.com/+biocomplexity



Research Symposium | 2016-11-01 | Abstracts

911 RESPONSE TIMES IN ARLINGTON COUNTY

Madison Arnsbarger* (Virginia Tech), Jessica Flynn* (Cornell U), Sam Tyner* (Iowa State), Gizem Korkmaz, Joshua Goldstein, and Vicki Lancaster (Social and Decision Analytics Lab)

*Data Science for the Public Good Student Fellows, Social and Decision Analytics Lab

Fire departments are under strict pressure to respond to emergency situations professionally, effectively, and efficiently. Some of the most important standards suggested by National Fire Protection Agency (NFPA) are related to the amount of time professional fire departments take to respond to emergency incidents. This is a particularly important standard to uphold because many studies have reported the remarkable relationship between response time and survival. An appropriate response time is not only vital to the general health and safety of citizens, but also the preservation of property.

In this study, we use Fire and Emergency Data from Arlington County over 6 years (2010-2015) with 870,000 observations and 86 different variables including the location and the time of the incident, CAD (computer-aided dispatch) call types, fire stations, responding units, and apparatus types. We define response time as the time between when a unit is dispatched and when the first unit arrives on the scene. We use a Bayesian linear model to analyze the effect of various predictors including hour of the day, CAD call type, station, year, month, and apparatus type, on response time to an incident ~1 mile away. We find that on average, response time increases steadily from midnight to 9:00 am, followed by a sharp decrease around 10:00 am. Incidents occurring in February (for all years) tend to experience the longest response times, while 2010 - 2015 show an increase with each year. We also analyze the response times by call type and the responding station. Our findings will be presented at the poster session.

STAY CONNECTED:

bivt.edu

facebook.com/biocomplexity

twitter.com/biocomplexityvt

youtube.com/c/biocomplexity

google.com/+biocomplexity