Book of Poster Abstracts

Thursday, May 29th, 2014

Hunter College of the City University of New York
Room 714 Hunter West Building
East 68th Street at Lexington Avenue, New York, NY 10065
EVENT SPONSORS

Hunter College of the City University of New York, Center for Translational and Basic Research
http://ctbr.hunter.cuny.edu

&

Clinical & Translational Science Center, Weill Cornell Medical College
http://www.med.cornell.edu/ctsc

Including its partners: Hunter College School of Nursing, Cornell University Cooperative Extension in NYC
Memorial Sloan-Kettering Cancer Center, Hospital for Special Surgery

Booklet Editor: Hilary T. Monaco
## Table of Contents

<table>
<thead>
<tr>
<th>Poster #</th>
<th>Poster Title</th>
<th>Page #</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Category: Basic</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Integration of metabolic and quorum sensing to achieve metabolic prudence in the regulation of bacterial virulence factors</td>
<td>5</td>
</tr>
<tr>
<td>2</td>
<td>Mathematical modeling of tumor-macrophage synergy and ecological therapy in glioblastoma</td>
<td>6</td>
</tr>
<tr>
<td>3</td>
<td>Pbx-directed control of cellular behaviors that drive face morphogenesis</td>
<td>7</td>
</tr>
<tr>
<td>4</td>
<td>Direct Arylation Via Microwaves In The Synthesis Of Aporphine-Inspired CNS Receptor Ligands</td>
<td>8</td>
</tr>
<tr>
<td>5</td>
<td>Temporal Dynamics of Positive Self-Referential Processing</td>
<td>9</td>
</tr>
<tr>
<td>6</td>
<td>Growth and Cooperation in Spatially Structured Communities</td>
<td>10</td>
</tr>
<tr>
<td>7</td>
<td>Electric Memories: Is Electric organ discharge (EOD) necessary for memory acquisition and consolidation?</td>
<td>11</td>
</tr>
<tr>
<td>8</td>
<td>Functional Studies of Nuclear Phosphorylated Forms of Tau and Possible Connections to Alzheimer’s Disease</td>
<td>12</td>
</tr>
<tr>
<td>9</td>
<td>Elucidating Cytokine Biomarkers in an Animal Model of Neuropsychiatric Disease</td>
<td>13</td>
</tr>
<tr>
<td><strong>Category: Bioinformatics</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Nup211 controls the expression of cytokinetic genes in fission yeast</td>
<td>14</td>
</tr>
<tr>
<td>11</td>
<td>Hadza Hunter Gatherer’s Sugar Intake: Implications for Western Health and Diet</td>
<td>15</td>
</tr>
<tr>
<td>12</td>
<td>Evolution in motion: emergence of multiflagellated and long-celled Pseudomonas aeruginosa in experimental evolution of swarming</td>
<td>16</td>
</tr>
<tr>
<td>13</td>
<td>Linking genotype to social phenotype in Pseudomonas aeruginosa</td>
<td>17</td>
</tr>
<tr>
<td>14</td>
<td>Multiscale Modeling of Functional Polymorphisms from Next Generation Sequencing Data</td>
<td>18</td>
</tr>
<tr>
<td>15</td>
<td>Building Capacity in HAI Prevention Research: NICHE and the STOP CAUTI Workgroup</td>
<td>19</td>
</tr>
<tr>
<td>16</td>
<td>Variability in gene expression provides insights on network topologies and function</td>
<td>20</td>
</tr>
<tr>
<td>17</td>
<td>MOLECULAR DRUG DESIGN FOR CHAGAS DISEASE</td>
<td>21</td>
</tr>
<tr>
<td>18</td>
<td>Computational approaches to identify DNA motifs for genes expressed in the mushroom body of Drosophila melanogaster</td>
<td>22</td>
</tr>
<tr>
<td>19</td>
<td>Computational Identification of Genotype-Specific Effective Drug Combinations</td>
<td>23</td>
</tr>
<tr>
<td>20</td>
<td>Computationally predicted off-targets explain the anti-cancer effects of metformin</td>
<td>24</td>
</tr>
<tr>
<td>21</td>
<td>Predicting Polypharmacology of Drug Metabolism and Transformation</td>
<td>25</td>
</tr>
<tr>
<td>22</td>
<td>DSSR, a program for Defining the Secondary Structures of RNA from three-dimensional coordinates</td>
<td>26</td>
</tr>
<tr>
<td>23</td>
<td>Computational Analyses of Nek-Family Proteins: Modeling, Evolutionary, and Mutational Studies</td>
<td>27</td>
</tr>
<tr>
<td>24</td>
<td>&quot;SPRING-DB&quot;: A Bioinformatics Pipeline for Automated Annotation of Pathogen Genomes</td>
<td>28</td>
</tr>
<tr>
<td>25</td>
<td>Integrative Analysis of Complex Cancer Genomics and Clinical Profiles Using the cBioPortal</td>
<td>29</td>
</tr>
<tr>
<td>26</td>
<td>Comparative meta-analysis of prognostic gene signatures for late-stage ovarian cancer</td>
<td>30</td>
</tr>
<tr>
<td>27</td>
<td>Using Bioinformatics to Determine which Proteolytic Enzymes cut GPI-Anchored Proteins into Peptides of Ideal Length for Mass Spec</td>
<td>31</td>
</tr>
<tr>
<td><strong>Category: Translational</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>A needs assessment among clinical providers of desired HIV information</td>
<td>32</td>
</tr>
</tbody>
</table>
SINaTRA: Species-INdependent TRANslation for Human Drugs and Disease

DREAM 9: Acute Myeloid Leukemia Outcome Prediction Challenge

Targeted sequencing and functional genomics in alopecia areata identifies ULBP6 as a critical node in its genetic architecture.

Resequencing of the T-cell costimulatory locus identifies variants carried on GWAS defined risk haplotypes in alopecia areata
INTEGRATION OF METABOLIC AND QUORUM SENSING TO ACHIEVE METABOLIC PRUDENCE IN THE REGULATION OF BACTERIAL VIRULENCE FACTORS

Kerry E. Boyle

Memorial Sloan Kettering Cancer Center and Weill Cornell

Pathogenic bacteria use density-dependent signaling (quorum sensing) to determine when they should cooperate as a population and when they should act as individuals. However, regulation of cooperative secretions is often influenced by nutrient conditions, suggesting that bacteria integrate information on nutrient levels with quorum sensing to regulate secretions and avoid exploitation by non-secreting mutants, a process called metabolic prudence. We investigated the synthesis of rhamnolipids, which are secreted virulence factors of the opportunistic pathogen Pseudomonas aeruginosa. By combining molecular biology and mathematical modeling techniques we characterized the regulatory dynamics of the rate-limiting enzyme for rhamnolipid production, RhlA. Growth curves with minimal media allowed us to probe how specific nutrient limitations influence growth and expression of rhlA. A mathematical model of the system enabled us to recapitulate bacterial growth and rhlA expression and make predictions about rhlA expression under different conditions and generate novel hypotheses to test experimentally. Together, our experimental data and mathematical model suggest that under growth limitation, the flux of carbon metabolites can be directed to rhamnolipid production without impacting the fitness of individual bacteria. In support of this hypothesis, we identified rhlA overexpression mutants with point mutations in genes in or related to the citric acid cycle. Rhamnolipids play a multifactorial role in the pathogenesis of P. aeruginosa, including biofilm dispersal and lysis of host immune cells. With a deeper understanding of the regulatory mechanisms conferring metabolic prudence we can attempt to influence P. aeruginosa to produce the ideal amount of rhamnolipids for a given situation, to disperse a biofilm or protect the host immune system.
Tumor-associated macrophages (TAMs) are known to be involved in the progression of many cancers, including glioblastoma. Experiments have shown TAMs in glioblastoma to be educated to a protumoral phenotype by glioma cell-expressed colony-stimulating factor 1 (CSF-1). Inhibition of TAM-expressed CSF-1 receptor (CSF-1R) by the molecule BLZ945 reduces the MRI volume of large glioblastomas in mice by reeducating TAMs to a neutral or antitumoral phenotype. The exact mechanism by which CSF1R inhibition perturbs the signaling between glioma cells and TAMs is unknown. We construct an ordinary differential equation model of educated TAM-driven glioblastoma growth and reduction under BLZ945 treatment. Our model predicts the blocking the recruitment of new macrophages from the periphery, in addition to TAM reeducation and inhibition of educated TAM protumoral signaling, to be necessary for the treatment outcome observed in vivo. We also predict a dependency of treatment outcome on the tumor state at time of treatment initiation. From these results, we conclude the depletion of peripheral macrophages observed in vivo under BLZ945 treatment and the tumor size at time of treatment to be important determinants of treatment efficacy.
**Pbx-directed control of cellular behaviors that drive face morphogenesis**

**Hila Chase**, Bingsi Li, PhD, Elisabetta Ferretti, PhD, Laura Quintana Rio, BSc, Sameer Khan, BSc, Licia Selleri, MD, PhD

Weill Cornell

Cleft lip/palate (CL/P) is the most prevalent human craniofacial birth defect world-wide (~1/700 live births). Affected individuals suffer from obvious facial deformation and are greatly impacted socially and communicatively. Despite its prevalence, not much is known about what causes CL. During development, fusion of nasal and maxillary processes at the lambdoidal junction (λ) occurs, requiring disintegration of the epithelial layer. CL occurs when the epithelial layer on each nasal process fails to disappear. Since in CP, Epithelial-Mesenchymal Transition (EMT, a process in which cells change fate) is known to have a role in fusion of the palatal shelves, we believe it may play a similar role during fusion at the λ, and hypothesize that impaired EMT may be a cause of CL. EMT is known to be regulated by Pbx transcription factors (TFs), which are also known to regulate cell proliferation, migration and apoptosis during facial morphogenesis. Pbx genes are expressed during fusion at the λ and their loss in Pbx compound mutant embryos (Pbx1−/−;Pbx2+/− and Pbx1−/−;Pbx3+/−) produces CL phenotype. To assess epithelial cell fates at the λ, we used cell lineage tracking experiments with reporter transgenic mice marking cells of epithelial origin at the λ. We used immunofluorescence to investigate EMT, apoptosis, and cell migration. To determine the role of Pbx TFs in those processes, we performed parallel experiments in wild-type and Pbx1/Pbx2 compound mutant embryos. The results obtained will help establish the role of Pbx-mediated EMT in the developmental causes of CL in mammals. Further analyses will also allow a better understanding of this disfiguring malformation.
DIRECT ARYLATION VIA MICROWAVES IN THE SYNTHESIS OF APORPHINE-INSPIRED CNS RECEPTOR LIGANDS

Juliel Espinosa, Nirav Kapadia and Wayne Harding PhD

Department of Chemistry, Hunter College of City University of New York, NY 10065

The 5-HT₂A receptor is known to play a significant role in various neuropsychiatric disorders such as schizophrenia and depression and blockade of this receptor is a promising avenue for the development of anti-drug abuse medications. We recently discovered that key alkyl or halogen modifications on aryl ring A (viz. C1 and C3 positions) of the aporphine alkaloid nantenine, affords significant improvements in 5-HT₂A antagonist potency. This revelation has prompted further modifications on the aporphine template as a means of delineating critical elements of the aporphine core scaffold that are required for antagonism of the 5-HT₂A receptor as well as selectivity vs other CNS receptors. Such a study has the potential to uncover new potent and selective 5-HT₂A antagonists as well as multi-receptor ligands that may be useful as biological tools.

Against this background, we designed and synthesized a series of C4 nantenine analogues that contain known 5-HT₂A antagonist pharmacophores embedded in the aporphine core structure. Our synthetic route here featured a Michael addition and a high yielding microwave-assisted direct arylation as key steps. In addition, we implemented an oxa-Pictet-Spengler cyclization and microwave-assisted direct arylation in the construction of a series of heterocyclic analogues in which the sole nitrogen atom was substituted by oxygen. Our work demonstrates the power and feasibility of microwave-assisted direct arylation in the synthesis of aporphine-inspired heterocyclic libraries and has allowed access to novel scaffolds for biological evaluation for the first time.
Temporal Dynamics of Positive Self-Referential Processing

Laura Fonseca, Emmanuel Garcia, Jean Quintero, Douglas Mennin

Background: Self-referential processing (SRP) refers to how individuals reflect on their own person, and can be indexed by examining their processing of personally relevant valenced stimuli. (Northoff et al., 2006). Emotion dysregulation in individuals with anxiety and depression has been associated with increased negative SRP (nSRP) and decreased positive SRP (pSRP), indicating a tendency to perseverate on negative reflections, rather than to reflect on experiences in such a way that increases or maintains positive affect (Mennin & Fresco, 2013). However, how this processing affects temporal dynamics of emotional responding has been little explored.

Methods: Healthy participants (n=25) passively viewed emotional images from the International Affective Picture System (IAPS; Lang et al., 2008) while connected to an electroencephalogram (EEG). Stimulus-locked event-related potentials (ERPs) were extracted from the raw EEG data and correlated with self-report questionnaire measures such as the Responses to Positive Affect questionnaire (Feldman et al., 2008) in order to evaluate the impact of pSRP on the time course of electrocortical responding to emotional stimuli.

Results: Electrode mean activations were examined for the late positive potential (LPP, occurring 400-1000ms post stimulus onset). A significant main effect for image type was found, with emotional images eliciting a greater mean activation than neutral images. Significant interactions were found between the emotional image type and high verses low levels of pSRP. The interactions were primarily driven by group difference on neutral versus mutilation images.

Conclusion: Preliminary results suggest that the LPP may reflect the impact of SRP on electrocortical response to emotion.
Growth and Cooperation in Spatially Structured Communities

Hilary T. Monaco, Kerry E. Boyle, Cong Huang, Joao B. Xavier

Memorial Sloan Kettering Cancer Center and Weill Cornell

Bacteria are highly social organisms that naturally exist in surface attached communities known as biofilms. Bacterial behavior in a biofilm is known to produce many shared products including virulence factors. We hypothesize that the nutrient environment experienced by bacteria within a biofilm is critical to many bacterial phenotypes including the expression of shared virulence factors. We have developed a model of *P. aeruginosa* populations and the nutrient environment they experience in liquid culture. This model is capable of accurately predicting cell growth and metabolically-prudent [1] expression of rhamnolipids, a communal good and a virulence factor. We are adapting this model to a spatially structured environment by describing the growth and virulence induction of bacterial colonies on a hard agar plates using a combination of mathematical and experimental techniques. We aim to understand this system such that we can control rhamnolipids production and gain a metabolic grasp of how *P. aeruginosa* participate in cooperative behavior in a spatially structured environment.

Electric Memories: Is Electric organ discharge (EOD) necessary for memory acquisition and consolidation?

Martha Ordonez

Hunter College

Previous studies have shown that weakly electric fish can find a goal in a complex maze. While learning the trajectory they generate a unique electric organ discharge pattern, EOD scalloping. The structural and functional homology between the mammalian hippocampus and the dorsal lateral gray mantle of teleost fish telencephalon in these fish has encouraged an investigation of PKMζ in Gnathonemus petersii, a weakly electric fish. PKMζ has a catalytic subunit making it constitutently active and also necessary for long-term memory. G. petersii were trained for 3 days with 6 trials/day and retested on day 4. Their telencephalon was removed, the dorsal lateral (DL) and dorsal medial pallium (DM) were excised and prepared for Western blot analysis. The results revealed and remarkable correlation between successful maze learning and the expression of PKMζ and GluA2. We conclude that the increased expression of these two molecular markers in the synaptic region is important for memory consolidation in G. petersii. Future work will design experiments to further strengthen the role of EOD scalloping in spatial learning and associated changes in PSD-95 and GluA2 markers.
Functional studies of nuclear phosphorylated forms of tau and possible connections to Alzheimer’s disease.

Syed Sarder, Frida E. Kleiman and Jorge Baquero

Hunter College

Neurons of Alzheimer’s disease patients show the development of paired helical filaments (PHF). These filaments are formed by insoluble hyperphosphorylated-tau aggregates. In normal individuals, tau is a highly soluble protein that stabilizes microtubules in the cytoplasm of neurons. Many efforts have been made to identify factors that might regulate tau phosphorylation under different conditions. For example, it has been shown that p73, a homologue of tumor suppressor p53, regulates the accumulation of phosphorylated-tau during Alzheimer’s disease. It was also shown that p44, another isoform of p53, promotes the phosphorylation of tau and activates the transcription of different tau kinases. Interestingly, Dr. Kleiman’s lab has recently shown that p53 can regulate mRNA 3’ processing in the nucleus during DNA damage response (DDR). As p53 and its isoforms are involved in nuclear functions during the cellular response to DDR; first we decided to examine the possibility that tau might localize in the nucleus of different cells. Our fractionation assays indicate that many tau isoforms are present exclusively in the nucleus of non-neuronal cells, such as HCT116 human colon carcinoma cell line. Interestingly, the pattern of these tau isoforms change after UV-treatment. We confirm the phosphorylation of nuclear tau by CIP treatment of nuclear extracts from these cells. Finally, we found that some of tau nuclear isoforms are ubiquitinated using in vivo ubiquitination assays. Although these are preliminary studies, these findings reveal a new possible connection between tau and other nuclear processes, such as DDR, mRNA 3’ processing and gene expression.
Elucidating Cytokine Biomarkers in an Animal Model of Neuropsychiatric Disease

Jeremy Seto

Hunter College

Human patients afflicted with neuropsychiatric illnesses have presented with significantly higher levels of circulating inflammatory cytokines independent of any infection. In a rodent model of schizophrenia, fetal mouse brains were analyzed to determine cytokine levels as a system of biomarkers for the disease state that may illuminate an underlying etiology from neurodevelopment alterations.
Nup211 controls the expression of cytokinetic genes in fission yeast

Ayisha R. Sookdeo, Hualin Zhong

Department of Biological Sciences, Hunter College; The Graduate Center, City University of New York

The Nuclear Pore Complexes (NPCs) are large multi-protein channels that traverse the nuclear envelope and mediate nucleo-cytoplasmic transport. Proteomic studies have revealed that NPCs are composed of about thirty different proteins called nucleoporins. In addition to their functions in nuclear transport, many nucleoporins play important regulatory roles in vital cellular processes, such as gene expression and cell cycle progression. The functions of numerous nucleoporins are still not clear. In our studies, we focus on a particular nucleoporin, nup211, in fission yeast. nup211 is an essential gene and is evolutionarily conserved from yeast to humans. The nup211 protein is localized on the nuclear side of the NPCs. Previous studies have shown that either overexpression or down-regulation of nup211 or its orthologs resulted in the nuclear accumulation of polyA-RNAs. We also found that altering expression of nup211 led to defects in cytokinesis including branched and elongated cells and cells with multiple septa. Furthermore, we observed that the expression of proteins important for cytokinesis, such as actin, was affected by nup211. To determine whether nup211’s function in cytokinesis is mediated via controlling gene expression at the transcriptional and/or mRNA export level, we are currently using RNA-Seq to analyze the profiles of mRNAs isolated from strains expressing different levels of nup211. The RNA-Seq data will shed light on the mechanism by which nup211 regulates cytokinesis.
Hadza Hunter Gatherer’s Sugar Intake: Implications for Western Health and Diet

Khalifa Stafford¹, Brian M. Wood², Jessica Rothman¹ and Herman Pontzer¹

¹City University of New York: Hunter College
²Yale University

Energy dense diets that are high in sugars are thought to be key contributors to the overweight, obesity, and diabetes epidemics. Traditional populations, in which the incidence of obesity and diabetes are low, provide an important point of comparison for understanding the causal factors underlying metabolic diseases. Here, we examined diet and obesity among the Hadza hunter-gatherers of Northern Tanzania. The Hadza, who maintain a lifestyle similar to our Pleistocene ancestors, consume 307 sugar calories, in the form of honey, per day. Based on empirical relationships, between sugar intake and prevalence of metabolic health conditions in 177 countries worldwide, the Hadza’s sugar intake would be expected to result in a 41.6% overweight occurrence, 12.6% obesity incidence, and a near 7.5% diabetes prevalence. To begin to assess why the Hadza fail to show a prevalence of these diseases, despite their large consumption of sugar, we determined the nutritional content in Hadza honey and compared the quantity of specific sugars with published U.S. honey data. Percentage of fructose, glucose, sucrose and maltose in Hadza honey were similar to the quantity of these sugars in U.S. honey. Our data suggests that other factors, such as having an active lifestyle, may play a role in preventing obesity and diabetes in the presence of a sugar-rich diet.

Dave van Ditmarsch & João B. Xavier

Memorial Sloan Kettering Cancer Center

Most bacteria in nature live in surface-associated communities rather than planktonic populations. Nonetheless, how surface-associated environments shape bacterial evolutionary adaptation remains poorly understood. We subjected *Pseudomonas aeruginosa* to repeated rounds of swarming, a collective form of surface migration. The population underwent remarkable parallel evolution to a hyperswarmer phenotype, where the characteristic branching of swarming motility changed into an almost circular morphology. The reproducible hyperswarming phenotype is always caused by parallel point mutations in a flagellar synthesis regulator, FleN, which make the naturally monoflagellated bacterium become multiflagellated. Hyperswarmers gain an advantage in swarming competitions, but they are strongly outcompeted in biofilm formation, which is an essential trait for *P. aeruginosa* in environmental and clinical settings. The finding that evolution in swarming colonies reliably produces evolution of poor biofilm formers supports the existence of an evolutionary trade-off between motility and biofilm formation. Furthermore, the mutations found in FleN oppose previously published phenotypes with knockouts of this flagella synthesis regulator, and it is of great interest to determine the molecular role of the point mutations. Preliminary computational analysis suggests that the mutated residues interact with FlhF, a known regulator of the number and polarity of the flagella. In addition, a remarkable clone evolved in a single population, which shows a drastic elongation (~3-fold) of the cell body. Cell elongation confers an additional advantage in swarming motility, seemingly without impacting other phenotypes.
Linking genotype to social phenotype in *Pseudomonas aeruginosa*

Jinyuan Yan

Memorial Sloan Kettering Cancer Center

The social phenotypes of *Pseudomonas aeruginosa*, such as biofilm formation, swarming motility and quorum sensing, are essential to its pathogenesis. Social phenotypes can be modulated through genomic alterations ranging from single point mutations, to complete gene deletions, and even large genomic rearrangements. However, the molecular mechanisms linking genotype and social phenotype remain largely unknown. We hypothesize that social phenotypes are main drivers of *P. aeruginosa* adaptation. We test our hypothesis using two complementary approaches: (1) experimental evolution and (2) investigation of *P. aeruginosa* genomic evolution using clinical isolates obtained from acute infections of cancer patients hospitalized at Memorial Sloan Kettering. Our long-term goal is to understand how social phenotypes of *P. aeruginosa* are shaped through selective pressures and their mechanisms of genomic regulation.
Multiscale Modeling of Functional Polymorphisms from Next Generation Sequencing Data

Li Xie¹, Clara Ng², Thahmina Ali³, Raoul Valencia³, Barbara L Ferreira³,⁴, Vincent Xue², Maliha Tanweer³, Dan Zhou⁵, Gabriel G Haddad⁵,⁶,⁷, Philip E Bourne¹,⁸*, Lei Xie²,⁹ *

¹Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California San Diego, La Jolla, CA 92093, USA.
²Department of Computer Science, Hunter College, The City University of New York, New York City, NY 10065, USA.
³Department of Biological Sciences, Hunter College
⁴Department of Psychology, Hunter College
⁵Department of Pediatrics, University of California, San Diego, La Jolla, CA 92093, USA.
⁶Department of Neuroscience, University of California
⁷Rady Children’s Hospital, San Diego, CA 92123, USA.
⁸Graduate Center, The City University of New York, New York City, NY 10016, USA.

One great challenge in modern biology class is to determine the functional roles of non-synonymous Single Nucleotide Polymorphisms (nsSNPs) on complex phenotypes. Statistical and machine learning techniques establish correlations between genotype and phenotype, but may fail to infer the biologically relevant mechanisms. The emerging paradigm of Network-based Association Studies aims to address this problem of statistical analysis. However, a mechanistic understanding of how individual molecular components work together in a system requires knowledge of molecular structures, and their interactions.

To address the challenge of understanding the genetic, molecular, and cellular basis of complex phenotypes, we have, for the first time, developed a structural systems biology approach for genome-wide multiscale modeling of nsSNPs, from the atomic details of molecular interactions to the emergent properties of biological networks. We apply our approach to determine the functional roles of nsSNPs associated with hypoxia tolerance in Drosophila melanogaster. The integrated view of the functional roles of nsSNP at both molecular and network levels allows us to identify driver mutations and their interactions (epistasis) in H, Rad51D, Ulp1, Wnt5, HDAC4, Sol, Dys, GalNAc-T2, and CG33714 genes, all of which are involved in the up-regulation of Notch and Gurken/EGFR signaling pathways. Moreover, we find that a large fraction of the driver mutations are neither located in conserved functional sites, nor responsible for structural stability, but rather regulate protein activity through allosteric transitions, protein-protein interactions, or protein-nucleic acid interactions.

Our studies demonstrate that the consolidation of statistical, structural, and network views of biomolecules and their interactions can provide new insight into the functional role of nsSNPs in Genome-Wide Association Studies, in a way that neither the knowledge of molecular structures nor biological networks alone could achieve. Thus, multiscale modeling of nsSNPs may prove to be a powerful tool for establishing the functional roles of sequence variants in a wide array of applications.
Building Capacity in HAI Prevention Research: NICHE and the STOP CAUTI Workgroup

Heidi Wald, MD, MSPH; Angela Richard, MS, RN; Brian Bandle, BA; Regina Fink, PhD, RN, AOCN, FAAN; Marie Boltz, PhD, RN; Sung-joon Min, PhD; Elizabeth Capezuti, PhD, RN

Hunter-Bellevue School of Nursing, Hunter College

Catheter-associated urinary tract infections (CAUTIs) are common among frail elders. CAUTI prevention relies on evidence-based nursing practices, few of which have been subject to multisite study. A cornerstone of prevention, surveillance (of indwelling urinary catheter (IUC) utilization and CAUTI rates) can be used to provide feedback performance improvement. Researchers at the University of Colorado (CU) partnered with the Nurses Improving Care of Healthsystem Elders (NICHE) program to create the STOP CAUTI Workgroup to implement and test the impact of electronic surveillance of IUC use and CAUTIs.

Development of the STOP CAUTI Workgroup was based on a modification of the Hopkins collaborative model, through a process to engage, educate, and establish an administrative framework prior to embarking on the study’s execute and evaluate phases. Recruited from among 245 NICHE member hospitals, 20 hospitals completed all steps required to participate in the cluster-randomized controlled trial of audit and feedback in the reduction of CAUTI among hospitalized patients. This poster details the engage, educate and establish stages of the project.
Variability in gene expression provides insights on network topologies and function

Daniel Carbajo

Albert Einstein College of Medicine

The study of gene expression variability represents a powerful approach to understanding biological systems that complements the typical view that accounts solely for population statistics like the mean. At a signaling pathway or regulatory network level, genes highly variable in their expression can confer higher plasticity to the network. These genes are thought of as being subject to flexible constraint in their expression, while lowly-variable ones are considered to be under more stringent constraint. An excessive variability, though, is a trait of complex diseases, like cancer. The analyses of variability profiles hence become relevant, as they may identify abnormal events that contribute to disease phenotypes. Inside a network, there is a relationship between gene expression variability and degree. Highly-connected genes at the core can potentially alter the expression of many genes while lowly-connected ones in the periphery could only alter the expression of few genes. We reasoned that highly-connected genes should be more tightly regulated, reflected by low variability in their expression.

We have studied the inter-individual gene expression variability profile of NMZL (nodal marginal zone lymphoma) patients, exploring potential unevenness of gene expression variability across networks (protein-protein interaction -PPI- and TF regulatory networks) to see how variability influences the resulting disease phenotype. A published dataset including copy number variation (CNV) and miRNA profiles was selected. No trend was found between CNV and variability, or between CNV and degree. Lowly-variable genes were found to be more highly connected, sitting at core positions in the networks and in significant pathways, while highly-variable ones were less connected, occupying peripheral positions. This differential positioning in the networks and pathways dictates distinct functional roles and cellular locations; while highly-constrained genes mainly operate in the nucleus and cytoplasm and are involved in basic functioning and maintenance of the cell, lowly-constrained genes are enriched for growth factor and receptor binding activities, located mainly at the plasma membrane or the extracellular space.

NMZL was compared to FL (follicular lymphoma), identifying differentially expressed and differentially variable genes, most of the latter being more variable in NMZL. These sets of genes are largely heterogeneous in the roles they play and the pathways and processes they are involved in, thus their distribution across the networks appears random. B-cell differentiation is tightly regulated by a miRNA-based program. We investigated whether differential regulation in the expression of miRNAs between the NMZL and FL was correlated to the differential variability observed for many genes between the two diseases. Such relationship, however, was not identified. This study provides insights on how gene expression variability affects a resulting disease phenotype, as it provides important information about network topologies; differential positioning of genes in the networks, depending on their expression variance, which could reflect the biochemical functions they perform in different locations of the cell.
Molecular Drug Design for Chagas Disease

Charles O Ogindo¹, Yayin Fang², Mozna H. Khraiwesh³, Clarence M. Lee³, Mohammad Ashraf⁴, William M. Southerland⁵ and Oladapo Bakare¹

¹Department of Chemistry, Howard University, 525 College Street, NW, Washington, DC 20059
²Department of Biochemistry and Molecular Biology, HU College of Medicine, Washington, DC 20059
³Department of Biology, HU, Washington, DC 20059
⁴Department of Comprehensive Science, Howard University, Washington, DC 20059

Chagas disease is an incurable disease of the heart and gastrointestinal tract with high mortality rates. It is caused by the protozoan parasite Trypanosoma cruzi (T. cruzi). More than 16 million people worldwide have been affected by Chagas disease. According to CDC only two drugs (Nifurtimox and Benznidazole) are available for treating Chagas disease and both of them have serious side effects. Those two drugs are only administered at the acute stage of the disease to bring it to chronic levels. Those critical problems with the current therapy limitations established the urgent need for fresh approaches to design and develop new drugs for Chagas disease. The presented research are to identify new drugs for Chagas disease with enhanced bioactivity and selectivity while having none or very limited side effects. Tubulin is an important protein essential for cell division and survival and has been used as a target for various diseases including cancer. Our laboratory recently has synthesized a group of small compounds that exhibited trypanosomal tubulin polymerization inhibition and a cororally anti trypanosomal activities. This new discovery opened a door for a fresh approach to the design and development of new drugs for Chagas disease by selectively attacking the T. cruzi tubulin without interfering the human tubulin. This poster presents our initial studies in understanding the binding pockets and affinities for this class of compounds. Using known structure of human tubulin as template, the homology model of T. cruzi tubulin was established and the interaction between the model of T. cruzi tubulin and our compounds were evaluated. we observed that the least bioactive ligand showed no special interaction in the binding pocket, whereas, the most bioactive ligands had at least one strong interaction presented in the binding pocket. These preliminary findings points us in the desired direction of finding agents that strongly interact with T. cruzi tubulin while having weak or even no interaction with the human tubulin.

Acknowledgment:
This work was partially supported by grant # G12 MD007597 from NIMHD, NIH to RCMI program at Howard University, HU Advance-It grant from NSF and Howard University College of Medicine Bridge Funds and Pilot Study Awards program. We also acknowledge NSF-MRI grant number CHE-1126533 and the NSF funded LSAMP program for support.
Computational approaches to identify DNA motifs for genes expressed in the mushroom body of Drosophila melanogaster

Raymond A. Ford¹, Carolina Guerra², Ming-Ying Leung¹,², Kyung-An Han³

The University of Texas at El Paso
¹Department of Mathematical Sciences
²Bioinformatics Program
³Department of Biological Sciences

Through the use of web scraping scripts we obtained data of interest from the HHMI Janelia Farm Research Campus and FlyBase to construct 6931 FASTA files containing the DNA sequence data for D. melanogaster. In an effort to enable accessing the data more efficiently, we developed a Tcl/Tk based GUI application (FlyTrap). The combination of web scraping and FlyTrap spared us the burden of having to visit approximately 21,000 web sites. The Multiple Em Motif Elicitation (MEME) tool was used to identify conserved sequence motifs in D. melanogaster DNA that may allow selective expression of genes in the mushroom body gamma lobe, the brain structure responsible for learning and memory. MEME returned the 100 motifs in the DNA segments driving gamma lobe expression, but only seven were selected based on an E-value less than 10. This information can be used to identify the enhancer elements for the genes that are expressed in the neurons under investigation facilitating the study of the mechanisms underlying learning and memory. With these results, we can further identify additional genes expressed in the gamma neurons, which mediate dopamine and octopamine signals. Abnormal dopamine functions are responsible for various neuropsychiatric disorders including ADHD, autism, schizophrenia, Parkinsons disease, and drug abuse/addiction to name a few. This study would ultimately help understand the underlying pathological mechanisms.
Computational Identification of Genotype-Specific Effective Drug Combinations

Kaitlyn Gayvert, BA\textsuperscript{1,2}, Olivier Elemento, PhD\textsuperscript{1,3}

\textsuperscript{1}Institute for Computational Biomedicine, Department of Physiology and Biophysics, Weill Cornell Medical College, New York, New York, United States
\textsuperscript{2}Tri-Institutional Training Program in Computational Biology and Medicine, New York, New York, United States
\textsuperscript{3}Institute for Precision Medicine, Weill Cornell Medical College, New York, New York, United States

Recent successes in cancer treatment can be attributed to the use of targeted therapies, which are designed to specifically target molecules that are involved in carcinogenesis. Although tumors are often sensitive to these treatments initially, drug resistance often develops and renders the applied drug ineffective. A promising alternative that has been successful in addressing this problem is through the use of combination therapies. However there are not enough resources to test all possible combinations through trial and error. Thus current studies are usually based on extremely detailed knowledge of drug mechanisms, which generally are not fully understood. To address this problem, we present a machine learning approach to identify effective drug combinations based on single drug efficacy information. We specifically built a model to predict effective combinations for mutant BRAF melanomas, although our approach is not specific to that case and is applicable to any cancer type. We trained the model with 496 effective and 986 non-effective combinations using the single-agent efficacies (GI50) in 15 mutant BRAF, 6 mutant RAS, and 6 wtBRAF/wtRAS cell lines. Using cross-validation, this model successfully achieves high accuracy (>0.85), sensitivity (~0.8), specificity (>0.9) and area under the ROC curve (>0.9). Overall, we present a broadly applicable approach for predicting effective drug combinations based on single-agent efficacies.
Computationally predicted off-targets explain the anti-cancer effects of metformin

Tom Hart
Hunter College

It is a grand challenge to identify biomarkers that determine the anti-cancer sensitivity of a specific drug for a specific patient due to high inter-tumor and intra-tumor heterogeneity. Although next-generation sequencing has generated a large amount of genomics data, conventional statistics techniques often fail to identify the functional roles of rare mutations that may drive drug responses. Here we introduce a novel structural systems biology approach to prioritizing critical genes and molecular interactions that are responsible for drug response phenotypes. By integrating context-specific data from structural genomics and functional genomics, and combining genome-wide protein structural modeling and biological network analysis, it is possible for us to identify biomarkers for personalized anti-cancer drug sensitivity. As a case study, we apply our methods to metformin, a drug that is widely prescribed to treat type 2 diabetes, but exhibits strong anti-cancer effects through activation of AMPK signaling. Through structural systems biology analysis, we predict that MAPK14, IL12B, AKT1, SEMA7A, and MAP2K2 genes are potential molecular targets of metformin. Furthermore, we suggest that several cancer-related genes, such as TP53, AKT1, PCNA, JUN, ESR1, and HNRNPK, represent the most important nodes within the biologically-true interaction network through which metformin elicits its anti-cancer effect. The identification of molecular targets and genetic factors that affect the pharmacology of metformin will shed new light on developing metformin as a safe, effective, personalized anti-cancer therapy.
Cytochrome P450 (CYP450) enzymes act on exogenous as well as endogenous substrates. These enzymes play important roles in drug metabolisms and transforming a prodrug into an active drug. The polymorphism of CYP450s accounts for the individual difference in drug responses. Thus information on how drugs interact with CYP450s will be critical for developing safe, effective, and personalized medicine, and recognizing drug-drug interactions. The multi-target binding profile (polypharmacology) of a large number of drugs with CYP450 remains unknown. Experimental screening is both time-consuming and expensive, and will benefit from computational predictions. However, few algorithms are able to reliably predict polypharmacology of CYP450-drug interactions. Here we apply novel multi-label machine learning algorithms and multiple chemical descriptors to predict the binding profile between five CYP450s and uncharacterized chemicals. In our preliminary studies, the best model achieves 61% of recall and 39% of precision. We will use case-based reasoning to further improve the model accuracy. It is expected that our model will provide a useful tool for pharmacogenomics and personalized medicine.
DSSR, a program for Defining the Secondary Structures of RNA from three-dimensional coordinates

Xiang-Jun Lu

Columbia University

As the number of experimentally solved RNA-containing structures grows, it is becoming increasingly important to characterize the geometric features of the molecules consistently and efficiently. Existing RNA bioinformatics tools are fragmented, and suffer in either scope or usability. DSSR, a new 3DNA program for Defining the Secondary Structures of RNA from three-dimensional coordinates, is designed to streamline the analysis of 3D RNA structures. It consolidates, refines, and significantly extends the functionality of 3DNA for RNA structural analysis.

Starting from an RNA structure in PDB or PDBx/mmCIF format, DSSR employs a set of simple geometric criteria to identify all existent base pairs (bp): either canonical Watson-Crick and wobble pairs or non-canonical pairs with at least one hydrogen bond. The latter pairs may include normal or modified bases, regardless of tautomeric or protonation state. The program denotes each bp by common names, the Saenger classification scheme of 28 H-bonding types, and the Leontis-Westhof nomenclature of 12 basic geometric classes.

DSSR detects multiplets (triplets or higher-order base associations) by searching horizontally in the plane of the associated bp for further H-bonding interactions. The program determines double-helical regions by exploring vertically in the neighborhood of selected bps for base-stacking interactions, regardless of backbone connection (e.g., coaxial stacking of helices). DSSR then identifies hairpin loops, bulges, internal loops, and multi-branch loops (junctions), and recognizes the existence of pseudo-knots. The program outputs RNA secondary structure in dot-bracket notation (dbn) and connect table (.ct) format that can be fed directly into visualization tools (such as VARNA).

DSSR classifies dinucleotide steps into the most common A-, B-, or Z-form double helices, calculates commonly used backbone torsion angles, and assigns the consensus RNA backbone suite names. The program also identifies A-minor interactions, ribose zippers, G quartets, kissing loops, U-turns, and kink-turns. Furthermore, it reports non-pairing interactions (H-bonding or base-stacking) between two nucleotides, and contacts involving phosphate groups.

Currently at version 1.1, DSSR is in a stable and mature state. A simple web interface and a comprehensive user manual are available. Supported by Dr. Robert Hanson, DSSR has recently been integrated into Jmol, a popular molecular graphics program. DSSR-related news and information can be found on the 3DNA homepage (x3dna.org). Questions and suggestions are always welcome on the 3DNA forum (forum.x3dna.org).
Computational Analyses of Nek-Family Proteins: Modeling, Evolutionary, and Mutational Studies

Dawid Plaza

CUNY Baccalaureate for Unique and Interdisciplinary Studies

Involvement of protein kinases, such as NIMA-like Kinases (Neks), in the cell cycle is linked with numerous cancerous phenotypes due to their variability of function, structure, and various interactions. Because of the potential of these proteins as chemotherapeutic targets, it is critical to understand the structure-function relationships of these proteins to ascertain their role in cancer. We have generated theoretical three-dimensional models of the Nek-family kinase catalytic domains, and elucidated structural and functional components critical to enzyme function through comparison with established Nek-family protein subfamily groupings based on phylogeny. Mutations from the Catalogue Of Somatic Mutations In Cancer (COSMIC) for the Nek-family proteins were evaluated for cancer driver characteristics using the CHASM algorithm. It was found that mutations in the critical regions of the enzyme were most likely to possess "cancer driving" character. Although further evaluation and rigorous statistical treatment will be used to quantify these data, our initial results suggest that structural domains housing the majority of these driver mutations are optimal targets for chemotherapeutic agents used to combat various Nek-linked cancers.
"SPRING-DB": A Bioinformatics Pipeline for Automated Annotation of Pathogen Genomes

Rayees Rahman¹, Raymond Liang¹, Jinyuan Yan², Joao Xavier² and Weigang Qiu¹

¹Department of Biology, Hunter College, City University of New York.
²Program in Computational Biology, Memorial Sloan-Kettering Cancer Center.

Pseudomonas aeruginosa is an opportunistic pathogen known for its environmental versatility and its ability to cause disease in immune compromised patients. Aside from the 11 published strains, new clinical strains are constantly being discovered displaying novel phenotypes that can exacerbate disease prognosis. Here we describe a bioinformatics pipeline tentatively named "SPRING-DB" (Scalable Pathogen Repository and INtegrative Genomics Database) to annotate, archive, and analyze sequence data obtained from a variety of sequencing platforms in order to characterize genes related to Pseudomonas pathogenesis. The pipeline includes the following key informatics components: (1) automated open reading frame (ORFs) discovery, (2) a relational database of P. aeruginosa genomes, (3) identification of orthologous genes, (4) reconstruction of genome phylogeny based on the “core genome”, which are the single-copy orthologs present in all sequenced genomes, and (5) a web-based genome browser. The core genomes were concatenated together to create a phylogenetic tree, which allows the visualization of newly sequenced strains relative to previously established strains and enables the comparative analysis of a variety of sequencing methodologies. Overall, the pipeline serves as a critical point in the work flow of genome analysis as it supports the pushing and pulling of genomic data in an automated and efficient manner. In the future, this pipeline will be expanded for use with SNP/haplotype identification, transcriptome analysis, gene linkage, or pan-genomic data.
Integrative Analysis of Complex Cancer Genomics and Clinical Profiles Using the cBioPortal

Yichao Sun

Memorial Sloan Kettering Cancer Center

The cBioPortal for Cancer Genomics (http://cbioportal.org) provides a web resource for exploring, visualizing, and analyzing multidimensional cancer genomics data. The portal reduces molecular profiling data from cancer tissues and cell lines into readily understandable genetic, epigenetic, gene expression, and proteomic events. The query interface combined with customized data storage enables researchers to interactively explore genetic alterations across samples, genes, and pathways and, when available in the underlying data, to link these to clinical outcomes. The portal provides graphical summaries of gene-level data from multiple platforms, network visualization and analysis, survival analysis, patient-centric queries, and software programmatic access. The intuitive Web interface of the portal makes complex cancer genomics profiles accessible to researchers and clinicians without requiring bioinformatics expertise, thus facilitating biological discoveries.
Comparative meta-analysis of prognostic gene signatures for late-stage ovarian cancer

Levi Waldron¹, Benjamin Haibe-Kains², Aedín C. Culhane³, Markus Riester³,⁴, Jie Ding³,⁴, Victoria Wang³,⁴, Svitlana Tyekucheva³,⁴, Christoph Bernau⁵, Thomas Risch⁷, Benjamin Ganzfried³, Curtis Huttenhower⁴, Michael Birrer⁶, and Giovanni Parmigiani³,⁴

¹City University of New York School of Public Health, Hunter College, New York, NY, USA
²Princess Margaret Cancer Centre, University Health Network, Toronto, ON, Canada
³Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute, Boston, MA, USA
⁴Department of Biostatistics, Harvard School of Public Health, Boston, MA, USA
⁵Institute for Medical Information Sciences, Biometry, and Epidemiology, LMU Munich, Germany
⁶Center for Cancer Research, Massachusetts General Hospital, Boston, MA, USA

Numerous gene signatures of patient prognosis for late-stage, high-grade serous ovarian cancer (HSOC) have been proposed, but diverse data and methods make these difficult to compare or prioritize for clinical application. However, the corresponding volume of publicly available expression data creates an opportunity to validate previous findings and to develop more robust signatures. To establish the current state of HSOC prognostic gene signatures, we undertook a systematic review and meta-analysis of the publicly available microarray data and published prognostic gene signatures. This effort generated a curated database of 1,319 clinically annotated HSOC microarray assays from 10 studies (Ganzfried et al., DATABASE 2013) and implementations of 14 published prognostic signatures for evaluation in independently available data. For each published prognostic signature we evaluated 1) its reproducibility, 2) its prognostic accuracy for overall survival in independent datasets, and 3) the prognostic value of the genes used in each signature relative to random gene sets. Twelve published models performed better than 97.5% of randomized risk scores, and six out-performed 97.5% of random signatures of the same size trained on the same data. The four top-ranked models achieved overall validation C-Indices of 0.56 to 0.60, and shared anti-correlation with expression of immune response pathways. Most models demonstrated lower accuracy in new datasets than in validation sets presented in their publication. This analysis provides definitive support for several prognostic models, but confirms that these require improvement to be of clinical value. This work addresses outstanding controversies in the ovarian cancer literature, including whether the prognostic signature proposed by The Cancer Genome Atlas Consortium (Nature 2011) improved on existing signatures, the impact of batch effects on a frequently re-used, high-profile dataset (Dressman et al., J Clin Oncol 2007), and the prognostic value of a recently proposed DNA damage repair signature (Kang et al., J Natl Cancer Inst 2012). Finally, this work provides a reproducible framework for meta-analytic evaluation of gene signatures.
Using Bioinformatics to determine which Enzymes cut Proteins into Polypeptide Sequences of an Ideal Length

Wurm, A.\textsuperscript{1,2,*}, Yazdi, S.\textsuperscript{2,3,*}, Aguilar, C.\textsuperscript{1,4,5}, Polanco, G.\textsuperscript{1}, Leung, M-Y.\textsuperscript{1,2,3,4}, Almeida, I.C.\textsuperscript{1,2}

\textsuperscript{1}Border Biomedical Research Center, Dept. of Biological Sciences,
\textsuperscript{2}Bioinformatics Program,
\textsuperscript{3}Dept. of Mathematical Sciences,
\textsuperscript{4}Computational Science Program, The University of Texas at El Paso, El Paso, Texas 79968, USA
\textsuperscript{5}Currently at the National Institute for Mathematical and Biological Synthesis, University of Tennessee Knoxville, Tennessee 37996-3410, USA

*These authors equally contributed to this study

The focus of this project is to determine which enzymes cut proteins into peptide sequences of ideal lengths for mass spectrometry (MS) analysis in order to better identify biomarkers for cancer detection. Glycosylphosphatidylinositol (GPI) is a posttranslational modification of proteins. GPI-anchored proteins (GPI-APs) can possibly be used to identify biomarkers for prostate cancer, a leading cause of cancer deaths in men. Studies have shown that GPI-TA, the enzyme that attaches the GPI anchor to the protein, is overexpressed in prostate cancer tissue. To prepare GPI-APs for MS, three enzymes – Asp-N Endopeptidase, LysC Lysyl Endopeptidase, and Trypsin – were chosen to cleave the proteins. Each enzyme follows a specific set of rules on where to cut a protein. We have written a series of Perl scripts that use the rules of each enzyme to virtually cut protein sequences into theoretical peptides in order to identify which enzyme, or combination thereof, will provide the greatest number of ideal length peptides. The ideal peptide length for detection by MS falls between six and twenty amino acids. We are further exploring the effect that the small probability to miss a cut at each possible site has on the choice of enzymes.
A needs assessment among clinical providers of desired HIV information

Michael Castro
Hunter College

OBJECTIVES: To understand the preferred content of HIV prevention/treatment information, and access mechanisms of medical providers and leaders from community based organizations (CBOs) who work with HIV diagnosed or HIV high risk people. METHODS: The Center for HIV Educational Studies and Training (CHEST) conducted an online survey of NYC-based providers and CBO leaders working in the HIV/AIDS area. CHEST used internally generated lists and a modified snowball technique.

RESULTS: Preliminary data show respondents (10 started and 7 complete) worked with populations on transmission methods that represented earlier stages of the HIV/AIDS epidemic but sought information across more diverse communities and were interested in issues of stigma and discrimination as well as behavioral based transmission. The internet is key to accessing new HIV information. Most respondents requested that new information be presented in short summary reports.

CONCLUSIONS: Should the preliminary findings hold up in a more extensive survey under development, CHEST should consider focusing on developing short summaries of past and ongoing studies with an emphasis on internet distribution as opposed to a concentration on publishing. Furthermore CHEST may consider even further broadening the scope of their research topics to ensure that all HIV populations and transmission factors are accounted for.
**SINaTRA: Species-INdependent TRAnsliteration for Human Drugs and Disease**

**Alexandra Jacunski**, MS\textsuperscript{1-4}; Scott J. Dixon, PhD\textsuperscript{5}; J.M. Silva\textsuperscript{3-6}; Brent R Stockwell, PhD\textsuperscript{3,6-8}; Nicholas P Tatonetti, PhD\textsuperscript{2,4}

Columbia University, New York, NY
\textsuperscript{1} Integrated Program in Cellular, Molecular, and Biomedical Studies
\textsuperscript{2} Department of Biomedical Informatics
\textsuperscript{3} Department of Systems Biology
\textsuperscript{4} Department of Medicine
\textsuperscript{5} Department of Biological Sciences
\textsuperscript{6} Department of Pathology and Cell Biology
\textsuperscript{7} Department of Chemistry
\textsuperscript{8} Howard Hughes Medical Institute

**Background:** Translating relationships from model systems to humans is a key challenge in biomedical research. Even well-studied properties, such as genetic interactions, are poor predictors of their human counterparts. We developed the SINaTRA (Species-INdependent TRAnslitation) algorithm to predict interactions that are difficult to study in humans by using normalization strategies to translate network parameters between species.

**Methods:** We constructed protein-protein interaction (PPI) networks for *S. cerevisiae*, *S. pombe*, and humans, then calculated the mathematical network parameters of all gene pairs in each network. Parameters were normalized within each species. We constructed a species-independent model of one genetic interaction, synthetic lethality (SL), using normalized network parameters and experimental SL data in *S. cerevisiae*. We validated the success of this model in *S. pombe*, then used the model trained on *S. cerevisiae* to predict SL in humans, providing all gene pairs with a SINaTRA score of SL between 0 and 1. We performed three validation experiments of our human results: two using published experimental SL screens, and one drug synergy screen using chemical inhibitors of three predicted SL pairs (BRAF/MAP2K1, BRAF/MAP2K2, and MAP2K1/MAP2K2) alone and in combination in HT1080 cells.

**Results:** Within-species SL models performed well in both *S. cerevisiae* and *S. pombe*, achieving AUCs of 0.92 in both species; there was not enough information on human SL to construct an appropriate model. A non-normalized model of SL in *S. cerevisiae* had poor performance when applied to *S. pombe*, achieving an AUC of only 0.63; in contrast, normalization increased interspecies translation success to an AUC of 0.86. Using our predictions of SL in humans, we found that validated KRAS-dependent and MYC-dependent SL pairs from two independent studies had significantly higher SINaTRA scores (p<0.0001 for both). Finally, our experimental validation showed synergy between BRAF/MAP2K inhibitors.

**Conclusion:** We successfully validated our species-independent model of SL by training a model in *S. cerevisiae* to predict SL in another species, *S. pombe*. We then predicted SL in humans by using this model to give all gene pairs a SINaTRA score. We validated these predictions using three experiments and, in so doing, created the first system-wide model of synthetic lethality in humans.
Acute Myeloid Leukemia (AML) is the most common acute leukemia affecting adults, accounting for close to 1.2% of cancer deaths in the U.S. This disease is made more challenging by possessing several subtypes, each of which may require different treatments and have different prognoses. Depending on subtype, five-year survival rates vary from 15–70%, and relapse rates vary from 33–78%. Better tools for determining subtype could aid in guiding treatment. The DREAM 9 AML Challenge presents to the crowd the task of using clinical data, cytogenetic markers and expression levels for 231 proteins (RPPA data) to generate predictive models of AML.

Participants will determine which models best predict whether AML patients will respond to ARA-C therapy, that is, experience Complete Remission (CR) or be Primary Resistant to therapy (PR). They will also predict remission duration for CR patients, as well as the overall survival for all AML patients. From this challenge, we expect to identify novel predictors that could not only improve AML patients’ outcome and survival but also lead to the discovery of potential biomarkers and therapeutic agents.

For details on how to participate please visit http://www.the-dream-project.org/. DREAM 9 will launch in Spring 2014.
Targeted sequencing and functional genomics in alopecia areata identifies ULBP6 as a critical node in its genetic architecture.

L Petukhova1,3, T Yamany3, EN Drill2,3, L. Bian3, Z. Dai3, M Duvic6, M Hordinsky7, D Norris8, V Price9, R Clynes3,4, and AM Christiano3,5

Columbia University
1Department of Epidemiology
2Department of Biostatistics
3Department of Dermatology
4Department of Medicine
5Department of Genetics and Development
6MD Anderson Cancer Center, Department of Dermatology, Houston, TX
7University of Minnesota, Department of Dermatology, Minneapolis, MN
8University of Colorado, Department of Dermatology, Denver, CO
9UCSF, Department of Dermatology, San Francisco, CA

Alopecia areata (AA) is a highly prevalent and poorly understood autoimmune disease which targets the hair follicle causing disfiguring hair loss. There is an enormous unmet medical need for the 5.3 million patients in the US who suffer from AA, arising primarily from a lack of understanding of disease pathogenesis. Our initial GWAS in AA revealed the first disease association to ULBP3/6 genes in any human disease. These genes are ligands for the NKG2D activating receptors for a repertoire of leukocytes. We biologically validated our statistical evidence by showing a marked upregulation of ULBP3/6 in lesional hair follicles and the presence of CD8+NKG2D+ T cells within the immune infiltrate. These findings, together with the previous demonstration of MICA overexpression in AA hair follicles, placed the NKG2D axis squarely at the center of AA pathogenesis, and invited a functional genomics approach to uncover causal variants predisposing to disease. In order to better understand the genetic variation driving the tagSNP associations identified in our GWAS, we selected a subset of 124 cases from our GWAS cohort for targeted deep resequencing with RainDance technology, amplifying 72Kb of sequence encompassing the entire region of association. As preliminary analysis of this dataset, we looked at the distribution of rare variants (p<0.01 in EVS and 1000G) across this region. We identified two rare missense variants, one of which is highly overrepresented in our cohort (p=0.005) and is located within ULBP6. Of the 127 rare or novel variants located within intergenic regions we identified 34 that fall within transcription factor binding sites, 7 of which are overrepresented in our cases, which cluster into two regions. One of these regions is a CTCF binding site, which is known to influence chromosome structure providing a mechanism for the regulation of gene expression. We have begun to assess the biological consequences of the ULBP6 protein coding variant, by developing a battery of functional assays aimed specifically at examining mRNA expression levels, protein levels, effects on cell surface display, receptor binding affinity and killing efficiency. We are additionally developing cellular assays to determine the effects of the regulatory variants. This work will clarify how GWAS identified genetic variation influences NKG2D-mediated cytotoxicity in the pathogenesis of autoimmune disease.
Resequencing of the T-cell costimulatory locus identifies variants carried on GWAS defined risk haplotypes in alopecia areata

Tarek Yamany¹, Lynn Petukhova¹,⁴, Esther N. Drill¹,⁵, Li Bian¹, Jane D. Broadbent³, Peter Nagy³, and Angela M. Christiano¹,²

College of Physicians & Surgeons, Columbia University, NY, NY
¹Department of Dermatology
²Department of Genetics and Development
³Department of Pathology and Cell Biology
Mailman School of Public Health, Columbia University, NY, NY
⁴Department of Epidemiology
⁵Department of Biostatistics

The costimulatory locus contains three key immunoregulatory genes (CD28, CTLA4, and ICOS), and was one of the first genomic regions associated with autoimmunity outside of the HLA locus. GWAS have provided robust agnostic evidence for association of this region with type 1 diabetes, rheumatoid arthritis, celiac disease, Graves disease and alopecia areata (AA). In our AA GWAS, we identified 8 typed and 59 imputed variants with statistically significant association, and conditional analysis on the typed SNPs identified at least two independent risk haplotypes at the locus. Despite the extensive evidence for conferring risk, the mechanisms by which genetic variants contribute to disease have remained elusive. In order to systematically identify all AA risk variants at this locus, we performed targeted resequencing of the costimulatory locus in 122 AA patients from our GWAS cohort. We targeted a 297Kb region including all exons and introns of CD28, CTLA4, ICOS, as well as all intergenic and flanking regions, utilizing RainDance technology and next generation sequencing. This experiment identified 1209 variants that passed rigorous QC filters, which we computationally phased, allowing us to assign alleles to chromosomes. Among the 244 sequenced chromosomes, we identified 88 chromosomes that carried either one or both of the GWAS-identified risk haplotypes. We next identified 208 common variants which were significantly enriched on the chromosomes carrying GWAS risk haplotypes (p<4.1e-5). Among these enriched SNPs, there is one CTLA4 protein coding variant (rs231775; p. T17A), five SNPs annotated with regulatory functions, and 17 SNPs that fall within transcription factor binding sites. We also identified 30 rare and enriched SNPs, 8 of which were identified to have regulatory functions. Targeted deep resequencing of GWAS loci has allowed us to fully enumerate all variants carried on the GWAS risk haplotypes, allowing us to move forward in identifying genetic contributions to complex disease.