3rd LA Conference on Computational Biology & Bioinformatics

Apr. 17 & 18, 2015

Center for Computation & Technology

LSU - ORED

- Bioinformatics
- Molecular Modeling
- NGS Sequencing
- Genomics
- Metagenomics

- Proteomics
- Cancer Informatics
- Infectious Diseases
- Drug Discovery
- Medical Diagnostics

LBRN - INBRE

CEIDR - COBRE

Keynote Speaker

Anton Nekrutenko (Penn St. U)
James Cavalcoli (Univ of Mich)

Christopher Mason (Weil Cornell Med.)
Michael J Salbaum (PBRC)

At LSU Digital Media Center

- **For registration, agenda, and other detail information, please visit Bioinformatics Conference page : http://lbrn.lsu.edu/events/bioinformatics-conference/**
- **Conference Registration and Hotel Lodging (if required) Deadline: Wednesday, April 1st, 2015**
  Poster Abstract Deadline: Wednesday, April 1st, 2015

Support from NIH/NIGMS grants 10P20GM103424 and 1P30GM110760 is gratefully acknowledged

3rd Annual Louisiana Conference on Computational Biology and Bioinformatics

April 17-18, 2015

Digital Media Center, LSU Baton Rouge

Agenda

April 17

1:10 pm - 1:20 pm……**Welcome and Opening Remarks**

Dr. Gus Kousoulas, Assoc. Vice Pres (LSU ORED) / Dr Thomas R Klei, PI LBRN

Session I:
Computational Tools and Bioinformatics I
Chair: Dr. Brygg Ulmer

1:20 pm - 2:20 pm……..….**Dr. Anton Nekrutenko**

"Next-generation sequencing data interpretation: enhancing reproducibility and accessibility"

Penn State University, PA

2:30 pm - 2:55 pm……..……**Dr. Brygg Ullmer**

"Toward computational genomics support via ecologies of tangible interfaces and interactive supercomputing"

Louisiana State University CCT

2:55 pm - 3:20 pm……..**Dr. Urska Cvek**

"Multidimensional visualization tools with the slider metaphor applied to a longitudinal drug study"

LSU Shreveport

3:20 pm - 3:45 pm……..**Dr. Diana Williams**

"The Transcriptome of Mycobacterium leprae During Log Phase Growth"

National Hansen's Disease Program (LSU-SVM)

3:45 pm - 4:00 pm……**Break**

Session II:

Computational Tools and Bioinformatics II

Chair: Joohyun Kim

4:00 pm - 5:00 pm……..**Dr. James Cavalcoli**

Bioinformatics Research Challenges: Producing high-quality, high-confidence & understandable results

University of Michigan Ann Arbor, MI

5:00 pm - 5:25 pm…… **Dr. Hye-Young Kim**

“Self-Assembled Nanostructures of VECAR molecules in Water: Molecular Dynamics Study”

Southeastern Louisiana University

5:25 pm - 5:50 pm……**...Dr. Seetharama Jois**

“Surfing PPI surfaces: Proteomimics and applications in drug design and imaging”

University of Louisiana at Monroe

5:50 pm - 6:15 pm……..……..……..**Dr. Surabhi Mahehwari**

“Molecular reconstruction of protein-protein interactions”

Louisiana State University at Baton Rouge

End of Day 1

Agenda

April 18

Session III: Applied Omics

Chair: Christopher Taylor

9:30 am - 10:30 am…….**Dr. Christopher Mason**

Integrative Genomics in Single Cells, Entire Cities, and Astronauts

Weill Cornell Medical College, NY

10:30 am - 10:55 am.…… **Dr. Kui Yang**

"Genome-wide screening of human haploid cell to identify cellular factor essential for herpes simplex virus replication"

LSU SVM Baton Rouge

10:55 am - 11:20 am..…… **Dr. Pushpendra Singh**

"Deciphering the molecular markers of mycobacterial transmission and evolution using comparative genomics"

National Hansen's Disease Program, LSU-SVM, Baton Rouge

11:20 am - 11:55 am ……..……..**Dr. John R. Caskey**

"Big Data and New Methods for Bioinformatics: Analysis from sequence to result"

LSU School of veterinary Medicine and LSU CCT

12:00 am - 1:30 am…… **Posters and Lunch**

Session IV: Infectious Diseases

Chair: James Cavalcoli

1:30 am - 2:30 pm…..……. **Dr. Michael J Salbaum**

"Functional Genomics of Birth Defects."

Pennington Biomedical Research Center, LA

2:30 pm - 2:55 pm…… **Dr. Juan J Martinez**

"Utilization of next generation sequencing technologies (RNAseq) to analyze the transcriptome of Rickettsia rickettsii during fatal Rocky Mountain Spotted Fever infections in a murine model of disease"

LSU-SVM, Baton Rouge

3:00 pm - 3:25 pm……..……**.Dr. Jong Hyun Ham**

"RNA-seq analyses revealed novel biological functions controlled by the intercellular signaling system of the rice pathogenic bacterium, Burkholderia glumae"

Louisiana State University-Baton Rouge

3:25 pm - 3:50 pm……..……**…..Dr. Rebecca Christofferson**

"Exploring arbovirus transmission assumptions"

LSU SVM Baton Rouge LSU Shreveport

3:50 pm ……**..Closing Remarks and Acknowledgements**

Abstracts for Oral Presentations

Friday April 17th, Session I

Computational Tools and Bioinformatics – I

Chair: Brygg Ulmer

"Next-generation sequencing data interpretation: enhancing reproducibility and accessibility"

1:20 – 2:20 pm

Anton Nekrutenko
Penn State University, PA

Areas of life sciences research that were previously distant from each other in ideology, analysis practices and toolkits, such as microbial ecology and personalized medicine, have all embraced techniques that rely on next-generation sequencing instruments. Yet the capacity to generate the data greatly outpaces our ability to analyse it. Existing sequencing technologies are more mature and accessible than the methodologies that are available for individual researchers to move, store, analyse and present data in a fashion that is transparent and reproducible. Here we discuss currently pressing issues with analysis, interpretation, reproducibility and accessibility of these data, and we present promising solutions and venture into potential future developments.

“Toward computational genomics support via ecologies of tangible interfaces and interactive supercomputing”

2:20 – 2:55 pm

Brygg Ullmer
Louisiana State University-Baton Rouge

Seven years ago, NIH Director Francis Collins is reported to have said 'It would not be good to have a $5,000 genome and a $500,000 analysis.' Today, this can be argued to be an approximation of the case. The present complexities and unknown future implications of such analyses loom large not only for scientists, but also for engagement by school children, senators, street people, solicitors -- each with increasingly complex, impactful nuances. We will describe two intersecting research threads leading to tangible genomics collaborations with several local, national, and international partners. These concern engagement in human and non-human genome consortium projects; and trajectories of synergistic research upon tangible and cyberphysical interfaces.

“Multidimensional visualization tools with the slider metaphor applied to a longitudinal drug study”

2:55 – 3:20 pm

Urska Cvek
LSU Shreveport

Considerable difficulty is encountered when evaluating drug-usage trends in a metropolitan setting; some methods utilize self-reporting, while longitudinal studies are often restricted to small cohorts. We instead obtained de-identified urological drug screening data from LSUHSC-Shreveport that covers a 13-year period (1998-2011) and consists of 111,359 visits by 71,311 patients. Though tests for specific drugs varied due to periodic changes in drug testing protocol, these tests consistently tested for amphetamines, barbiturates, benzodiazepines, cannabinoids, cocaine, and opiates. We constructed demographic profiles based on patient's ethnicity, gender, and age classification to determine detection patterns across each class. The majority of visits to LSUHSC-Shreveport include the African American (AA) and Caucasian (Cc) samples; AA represented the plurality of patients. Most notably, we observed a striking increase of positive tests for cannabinoids in the 0-4 age group; ~10-fold and ~5-fold increases were observed for this age group for AA and Cc respectively. Cocaine positive tests were predominantly found in AA children and adults; however, cocaine positives decreased during this period. Tests for both benzodiazepines and opiates were higher in Cc regardless of age; although benzodiazepine positives decreased in the general population over the past 10 years, this number had increased ~3-fold for both AA and Cc in the 0-4 age group, while opiate positives increased ~3-fold over the 13 year period (with greater positives in Cc). We were also able to use this data to construct a drug history of some patients, finding that 7,721 patients had a positive test on at least two separate visits. Of these patients, cannabinoids, cocaine, and benzodiazepines were most likely to yield a positive result (25.5%, 17.6%, and 16.2% respectively). Work is ongoing to determine this data's applicability respecting the evaluation of the gateway hypothesis.

“The Transcriptome of Mycobacterium leprae During Log Phase Growth”

3:20 – 3:45 pm

Diana Williams
DHHS/HRSA/HSB/National Hansen's Disease Programs

Mycobacterium leprae, the etiologic agent of leprosy, has undergone reductive evolution resulting in a genome with less than 50% protein coding capacity of its close relative, M. tuberculosis. Many genes and associated functions were lost possibly due to a drastic change in lifestyle from free-living, to host-dependent. To determine the functional genome of M. leprae, global gene expression was studied during log phase of growth in the immunocompromised mouse foot pad at 5 mths post infection (MFPML5) using RNA-Seq. These data were compared to that of M. leprae during short-term metabolic maintenance of bacteria in mycobacterial medium for 96 hr to potentially identify nutritional deficiencies of this medium. RNAs were prepared and sequenced using the SOLiDâ„¢ 5500 System (Genomics Core Lab, Pennington Biomedical Research Institute, BR, LA). Trimmed reads were mapped to the Mycobacterium leprae TN genome (Maverix Biomics, Inc, San Mateo, CA). Read counts were normalized across all samples and then used for differential expression analysis using DEseq. Averaged DEseq abundance levels of annotated reads demonstrated that 100% of 1614 CDSs, representing 10 functional groups, 85% of 1310 pseudogenes, and all 50 stable RNAs were expressed in MLMFP5. Significant differentially expressed genes were determined by adjusted P-value with a threshold of 0.05. Log2 (fold change) between samples. 37% of CDSs had at least 2-fold altered gene expression levels when compared to axenic held M. leprae. Approximately 60% of Cell Wall and Cell Processes group gene transcripts were significantly up-regulated in MLMFP5 including those encoding secretory systems and sugar transporters. Transcripts were detected for genes within the glycolytic, lipid and folate metabolic pathways of MLMFP5. Taken together these data demonstrate that M. leprae utilizes glucose and lipids as carbohydrate sources and is capable of producing folate within its intracellular macrophage niche. These data suggest that M. leprae utilizes its entire repertoire of protein coding genes for survival and growth in the MFP model giving it the capability to utilize host-derived metabolites. Conversely, several of these genes were down-regulated in axenic medium giving us potential clues to nutritional deficiencies in the current mycobacterial medium.

Friday April 17th, Session II

Computational Tools and Bioinformatics – II

Chair: Joohyun Kim

"Bioinformatics Research Challenges: Producing high-quality, high-confidence & understandable results"

4:00 – 5:00 pm

James Cavalcoli
University of Michigan

There is a rapid increase in the volume of sequence data is being generated at a reduced cost per base while at the same time, the bioinformatics analysis is getting more complex and the number of methods and effort needed are increasing. New methods and algorithms require testing and validation with gold-standard datasets to insure accurate, reproducible, high-confidence results in any lab or core facility. Testing different algorithms to identify the best method for the type of data produced is also a challenge. How do we get the highest quality results (e.g. sensitivity, specificity and lowest FP/FN rates). Lower sequencing costs results in an increase in the number of samples, and consequently an increase in the need to integrate larger datasets within and across experiments and. These larger collections of results and data require methods for visualizing and summarizing so that the results are understandable and interpretable by the researchers. Finally, experimental data needs to be annotated and linked to existing knowledge (e.g. TCGA, dbSNP, 1KGenomes, Literature) in order for scientists to be able to put their results into meaningful context. This talk will present some of the solutions we've developed to address these challenges.

“Self-Assembled Nanostructures of VECAR molecules in Water: Molecular Dynamics Study”

5:00 – 5:25 pm

Hye-Young Kim
Southeastern Louisiana University

VECAR is a newly synthesized molecule [1], which is an amphiphilic antioxidant molecule that consists of two molecular groups, vitamin-E and Carnosine, linked by a hydrocarbon chain. The hydrocarbon chain is hydrophobic and both vitamin-E and Carnosine ends are hydrophilic. In the synthesis process, the length of the hydrophobic chain of VECAR molecules can vary from the shortest (n=0) to the longest (n=18), where n indicates the number of carbon atoms in the chain. The self-assembled nanostructures (SANs) were expected to be used as a drug delivery system. We conducted a series of atomistic Molecular Dynamics simulations to study the self-assemblies of VECAR molecules of varying length in water using GROMACS on LONI HPC resources. Our study shows that there is a strong correlation between the shape and atomistic structure of the SANs and the chain-length (n) of VECAR molecules. We will report the methods and results of our study. [1] C. E. Astete, D. S. Meador, D. Spivak, C. Sabliov, Synthetic Commun. 43, 1299 (2013).

“Surfing PPI surfaces: Proteomimics and applications in drug design and imaging”

5:25 – 5:50 pm

Seetharama Jois
University of Louisiana at Monroe

Proteins interact with one another in an obligatory fashion maintaining a stable interaction for a long period of time, or interactions that are transient. These interactions control many biochemical pathways. A detailed knowledge about the structure of the interaction surface of proteins and its energetics is necessary to understand the regulatory mechanism of the biochemical pathway with the ultimate goal of modulating or blocking the biochemical pathways for therapeutic purposes. In comparison to protein-small molecule interaction, protein-protein interaction (PPI) surfaces are relatively large. In addition to this, surfaces of PPI are generally flat and have shallow grooves and pockets that are present at the surfaces of proteins that can bind to small molecules. Although PPI surfaces are large, some small regions of the interacting surfaces contribute to the binding energy more than the other regions. The regions on PPI interfaces that contribute more to the binding energy are called hot-spots. The presence of these hot-spots provides opportunity for targeting PPI with therapeutic agents because compounds that are designed to interact with hot-spots should prevent or block PPI. Interface peptides that mimic the proteins can be used to inhibit PPI, and the molecules designed based on interface peptides are called 'proteomimics'. Computational docking methods are used in several stages during the design of PPI inhibitors including docking of two proteins, finding hot-spots, and screening for a target compound. Using experimental and computational studies we have demonstrated that proteomimics can be designed from PPI interfaces and these can be used as a therapeutic agent or imaging agents. Two sets of PPIs, CD2-CD58 that are important in immunomodulation and EGFR that have significant impact on cancer therapy will be discussed. Funding for this research was provided by NIGMS/NIH grant 8P20GM103424 and 1R01 CA179902-01A1 (Dr. Vicente).

“Molecular reconstruction of protein-protein interactions.”

5:50 – 6:15 pm

Surabhi Mahehwari
Louisiana State University-Baton Rouge

Protein-protein interactions (PPI) mediate several biological processes at the molecular level. Thus,building three-dimensional PPI networks is important to interpret the information encoded in genomes. Predicting PPI sites and the structure of a protein complex are two important related components of this problem. Several computational protein-protein prediction methods have been developed in the past. However, majority of the existing methodologies are designed for experimentally determined protein structures. Because a large number of proteins in a genome will only have structure models available, computational tools must be tolerant to structural inaccuracies in order to be used for genome-wide modeling of PPIs. We contribute to this topic by proposing eFindSitePPI, a software for PPI prediction that capitalizes on the tendency of the location of binding sites to be highly conserved across evolutionarily related protein dimers. We show that eFindSitePPI is highly tolerant to structural inaccuracies in the query proteins and performs better for protein-models when compared to nine other state-of-the-art prediction methods. Furthermore, we developed eRankPPI, an algorithm to identify correct docking conformations of protein-models as well as experimental structures produced by docking softwares. The scoring function of eRankPPI uses several features including predicted interfaces with probability estimates calculated by eFindSitePPI and contact-based symmetry scores. A comparative study between eRankPPI and other state-of-the-art scoring methods shows that eRankPPI improves the success rate by ~10% on the benchmark dataset of homo and hetero complexes. The encouraging results obtained especially for protein-models open up the possibility for large scale reconstruction of structure-based PPI networks.

Saturday April 18th, Session III

Applied Omics

Chair: Dr Christopher Taylor

"Integrative Genomics in Single Cells, Entire Cities, and Astronauts"

9:30 – 10:30 am

Christopher Mason
Pennsylvania State University

The avalanche of easy-to-create genomics data has impacted almost all areas of medicine and science, and here we report the implementation of genomics technologies from the single-cell to an entire city, as well as integrative genomics approaches to space medicine. Recent methods and algorithms enable single-cell and clonal resolution of phenotypes as they evolve, both in normal and diseased tissues. Notably, some of these changes can be discovered by single-cell analysis and enable prognostic relevance. We also show that the genome, epigenome, transcriptome, and epitranscriptome all harbor some evidence of tumor evolution. Finally, we will discuss pilot data for creating enabling patients to become more involved in their 'omics data, including to an integrative genomics view of an entire city (based on our Pathomap project) that leverages longitudinal genomics and microbiome profiles of the NYC subway system. All of these pieces work together to guide the most comprehensive, longitudinal, mutli-omic view of human physiology with the NASA Twins Study, which launches on March 28th to the International Space Station (ISS) and enabling the most in-depth physiological and medical profile of a human being ever created.

“Genome-wide screening of human haploid cell to identify cellular factor essential for herpes simplex virus replication”

10:30 -10:55 am

Kui Yang
Louisiana State University-Baton Rouge

Herpesviruses depend on host proteins to enter cells, replicate their genome, and produce infectious progeny. Consequently, while herpesvirus and host cell interactions have been investigated extensively many have defied molecular identification. To identify cellular factors essential for herpes simplex virus 1 (HSV-1) replication, we performed genome-wide screening in the human haploid cell line HAP1. We generated a library of HAP 1 cells lacking one or more nonessential genes by inserting a retroviral gene-trap vector randomly into the cellular genome. By infecting these mutant cells with HSV-1, we identified 10 herpesvirus-resistant cell clones. HSV-1 replication in these cell lines was reduced from 1000 to 10,000 fold compared with those in wild type HAP1 cells. Through immunoblotting and electronic microscopy examination, it was determined that viral gene expression was impaired in some of these cell lines, while viral particles accumulated in either the nucleus or cytoplasm of infected cells in others. Through Next generation sequencing on the Illumina Hiseq2000 sequencer followed by bioinformatics analysis, we chose thirty host genes with high rate of gene-trap vector insertion to investigate the correlation of these genes with the resistant phenotypes. To verify the most relevant mutations, we plan to mutate individual candidate genes using the CRISPR Cas9 gene knock out system, generate individual clones of these cells, and then test for viral resistance in the resulting cell lines.

“Big Data and New Methods for Bioinformatics: Analysis from sequence to result”

10:55 – 11:20 am

John Caskey
Louisiana State University-Baton Rouge

Bioinformatics is a rapidly growing field that has barely kept pace with even more rapidly growing technology. As the capabilities of hardware have increased, the potential for Bioinformatics research tools has as well, most recently as advances in Metagenomics and Microbiome research. Bioinformatics tools are designed to accomplish many tasks, but at the most basic level, their purpose is to process millions of data points, and output a human-readable result, such as gene differentiation or variant calling. This result can either comprise an experiment in itself, or be an integral part of a larger scientific study. A need to scale analyses has arisen in recent years as the size of datasets has increased to a size beyond the capabilities of single machines, which has ushered in a set of frameworks to accommodate these vast analyses, commonly called Big Data. To demonstrate the capabilities of each setup, namely a single node, Big Data, and emerging technologies, a 'start to finish' analysis will be performed on a data set. The speed, efficiency, and accuracy will be compared to assess the role of each in the years to come.

“Deciphering the molecular markers of mycobacterial transmission and evolution using comparative genomics”

11:20 – 11:55 am

Pushpendra Singh
National Hansen's Disease Program, LSU-SVM, Baton Rouge, LA

Leprosy is a chronic disease caused by uncultivable pathogen Mycobacterium leprae and a newly described species M. lepromatosis. Over 220,000 new leprosy cases are recorded worldwide annually, nearly 60% of which are from India. The dynamics of M. leprae transmission and mechanisms by which it causes nerve damage remain largely unknown. Using whole-genome sequencing and comparison, we developed a reliable SNP-genotyping scheme that shows excellent association with geographic origins of M. leprae. (Nature Genetics 2009). We also investigated the natural leprosy infection in wild population of armadillos in southern US. These small mammals are high susceptibility to leprosy. Surprisingly, the strains present in armadillos were nearly indistinguishable from the human strains from that region. This was the first molecular evidence showing zoonotic link between armadillo and with human-leprosy (NEJM 2011). Subsequently, we used a hybridization-capture approach for genome-wide analysis of M. leprae directly from patient biopsies as well as from ~1000 year old leprosy skeletons from Europe. To our surprise, the comparative genomics revealed striking genome-level identity between ancient strains from UK and the current zoonotic strains present in Southern US, confirming the European origin of leprosy in Americas (Science 2013). Recently, we used innovative and inexpensive methods for DNA-enrichment using biotinylated baits to remove host DNA (Negative selection) and pathogen DNA (Positive selection). Using the combination of these approaches, we reported the first-ever genome of a new uncultivable pathogen M. lepromatosis which is associated with a severe form of human leprosy called 'diffuse lepromatous leprosy' and also identified in the Red-squirrels from Scotland. Comparative analysis of this newly identified species with M. leprae revealed that both species have undergone reductive evolution together and diverged around 13.9 million years ago (PNAS 2015). Further, this work provides valuable genomic insights into biology, pathogenesis and neuropathogenic potential of leprosy bacilli in particular and illuminates on the mycobacterial evolution in general. Cited refs: 1. Monot et al, Nature Genetics 2009; 41(12): 1282-9. 2. Truman et al, NEJM 2011. 364(17): 1626-1633. 3. Schuenemann et al, Science 2013; 341(6142): 179-83. 4. Singh et al, PNAS USA (in press, #2014-21504).

Saturday April 18th, Session IV

Infectious Diseases

Chair: Dr James Cavalcoli

"Functional Genomics of Birth Defects."

1:30 – 2:30 pm

J. Michael Salbaum
Pennington Biomedical Research Center

Maternal diabetes during pregnancy constitutes a major risk factor for birth defects: congenital malformations such as heart defects and neural tube defects are up to 10-fold more frequent. Furthermore, exposure to maternal diabetes during development can also cause long-term health consequences in the offspring, in a process termed 'developmental programming'. Despite many efforts, a unified and systematic understanding of the underlying pathogenic mechanisms remains elusive. To address these questions, we have focused on functional genomics and epigenomics of the developing embryo in response to diabetes exposure. Using mouse models of maternal diabetes and microarray technology, we have shown that maternal diabetes has profound effects on gene expression in the developing embryo by altering expression of transcription factors and chromatin modifiers. Maternal diabetes during pregnancy also increases the overall variation of gene expression levels in the embryo, suggesting a loss of regulatory precision. Hierarchical clustering analyses show that expression profiles distinguish exposed from non-exposed embryos, but cannot identify embryos with birth defects from the diabetes-exposed but unaffected group; however, using variability of gene expression as the test parameter provides the necessary resolution. Epigenetic modifications of DNA or chromatin are intricately involved in the regulation of gene expression, and as such are necessary for proper development. Using ChIPSeq technology, we can demonstrate that changes of H3K27 acetylation marks are significantly enriched (i) near genes known to cause neural tube defects in mouse mutants, and (ii) near critical regulators such as Ep300, suggesting that changes in histone acetylation patterns shape embryonic transcriptome responses to maternal diabetes exposure. To define the cellular basis for dysmorphologies, we used refined transcriptome analyses by combining laser microdissection of the open neural plate, SAGE next-generation sequencing, and systems biology approaches. This strategy allowed us to identify the cell population at the center of malformations, and a unified mechanism that explains birth defect phenotypes in diabetic pregnancies.

“Utilization of next generation sequencing technologies (RNAseq) to analyze the transcriptome of Rickettsia rickettsii during fatal Rocky Mountain Spotted Fever infections in a murine model of disease”

2:30 – 2:55 pm

Juan Martinez
Louisiana State University School of Veterinary Medicine, Department of Pathobiological Sciences

Rickettsial species are responsible for severe human diseases including re-emerging Rocky Mountain Spotted Fever (RMSF) infections in North America. Several rickettsial species have been utilized as surrogates to understand the pathology and pathogenesis elicited by spotted fever group (SFG) rickettsial species in animal models of infection. However, a bona fide animal model of disseminated fatal RMSF had not been described. We developed and characterized a model of disseminated, fatal RMSF via intravenous (i.v.) inoculation of Rickettsia rickettsii 'Sheila Smith' into C3H/HeN mice. Similar to the progression to fatal outcomes in human patients, mice developed severe lesions to the spleen and liver including neutrophilic splenitis, pyogranulomatous and necrotizing hepatitis, which correlated with a higher burden of rickettsiae. We hypothesized that R. rickettsii encounters diverse environments in vivo compared to those encountered in vitro and that rickettsiae respond accordingly to promote colonization and dissemination to target organs and tissues. We isolated total RNAs from purified bacteria and from R. rickettsii infected murine liver and spleen and then constructed the cDNA libraries for analysis on the Ion Protonâ„¢ System semiconductor sequencing platform. Using these technologies, we determined that R. rickettsii initiate transcriptional programs in vivo that are significantly different from those found during in vitro growth and include the modification/overexpression of lipopolysaccharide (LPS) and peptidoglycan biosynthesis, initiation of the SOS stress response, induction of genes involved in a type IV secretion system, and the remodeling of outer-membrane protein content. These results demonstrate that up-regulation and down-regulation of a distinct cohort of genes in vivo may contribute to the ability of R. rickettsii and other related rickettsial species to initiate successful infections and ultimately cause disease in infected mammals.

“RNA-seq analyses revealed novel biological functions controlled by the intercellular signaling system of the rice pathogenic bacterium, Burkholderia glumae”

2:55 – 3:20 pm

Jong Hyun Ham
Louisiana State University-Baton Rouge

Burkholderia glumae is the main causal pathogen for bacterial panicle blight of rice. The phytotoxin, toxoflavin, is a major virulence factor of this pathogen and its production as well as the virulence of the pathogen is regulated by intercellular communication, which is represented by the bacterial quorum-sensing (QS) dependent on the LuxI/LuxR homologs, TofI/TofR. In TofI/TofR QS, N-octanoyl homoserine lactone, an acyl-homoserine lactone (AHL)-type autoinducer (bacterial hormone), is produced by the TofI AHL synthase and recognized by the TofR transcriptional regulator. We recently identified another signaling component, TepR, which is homologous to LuxO proteins of Vibrio spp. In contrast to TofI/TofR QS, TepR negatively regulates the production of toxoflavin and the overall bacterial virulence, suggesting that counterbalancing interaction between TofI/TofR QS and TepR may be a crucial mechanism for modulating the intercellular communication of B. glumae. Recent comparative analyses of the transcriptome profiles from B. glumae strains having different phenotypes in the function of TofI/TofR QS and TepR suggest that Type VI secretion, heat shock protein production, and R-body formation are other biological functions controlled by these signaling components, in addition to toxoflavin production. This indicates that transcriptome analysis using RNA-seq approach can provide great insights into the comprehensive biological functions of a given signaling/regulatory factor.

“Exploring arbovirus transmission assumptions”

3:20 – 3:55 pm

Rebecca Christofferson
Louisiana State University-Baton Rouge

Transmission is a dynamic process consisting of so many working parts. Disassembly of this process leads to better understanding of the nuances of transmission in the context of major events, namely introduction and emergence, persistence and expansion, seasonality, and ultimately mitigation and control practices. In arbovirology where transmission necessarily involves an arthropod host, quantification and modeling of transmission is riddled with assumptions, which can lead to an underestimation of the probability of emergence and the inappropriate parameterization of models. To tackle these assumptions and offer better insight into transmission, I integrate laboratory, field and quantitative methods, focusing mainly on dengue and chikungunya viruses. My intention is to continue to bridge the fields of laboratory-based basic science with field data and relevancy and provide more insight into the parameterization of transmission models.

Abstracts for Posters

“Maintenance of Multiple Nef Functions Limits and Dictates Pathways of Immune Escape”

Nick Maness
Tulane University
PosterID: 1

CD8+ T lymphocytes (CD8TL) can effectively control SIV and HIV replication. CD8TL targeting of particular regions in the Nef globular core is associated with control of both viruses suggesting common mechanisms of control. Deep sequencing of SIV from nine Mamu-B*017:01+ (an MHC-I allele associated with enhanced control of SIV) macaques revealed distinct selection patterns in two Nef epitopes Acute mutations in the Nef195-203MW9 epitope impacted Nef's ability to downregulate Tetherin and CD28, both of which are modulated via Nef interactions with host clathrin adaptor proteins, with mutations impacting MHC-I anchor residues being most destructive. Mutations that impacted Nef's ability to downregulate Tetherin also led to dramatic increases in the amount of viral Envelope on the surface of infected cells, which is known to increase the susceptibility of these cells to antibody dependent cellular cytotoxicity. Thus, we may have discovered an unexpected link between T cell, antibody and NK cell immunity wherein CD8TL select for viral variants that are more vulnerable to ADCC. We are now investigating this novel finding. These data suggest mechanisms by which Nef targeting CD8TL can contribute to viral control; by targeting conserved epitopes involved in functions that are critical for high-level viral replication in vivo. Our immuno-virological approach to understanding T cell mediated control of SIV revealed novel insights into Nef structure:function relationships and suggest that particular regions of the Nef protein might be attractive components of vaccines designed to induce potent CD8TL responses.

“Identifying novel small molecules for improved antifungal drug treatment. ”

Kevin Murphy
Broad Institute, Koch Institute for Integrated Cancer Research at MIT, Division of Genetics, Department of Medicine, Brigham and Women's Hospital & Harvard Medical School
PosterID: 2

Invasive fungal infections (IFIs) are associated with high rates of morbidity and mortality and pose a serious health concern for severely immunocompromised patients. Fungal resistance to current drug therapies is largely due to the transcriptional upregulation of membrane associated efflux pumps. The fungal-specific zinc cluster (Zn2Cys6) family of transcription factors (TFs) are primarily responsible for the upregulation of these efflux pumps and thereby mediating pleiotropic drug resistance (PDR) in yeast. Thus, these TFs offer an attractive and rational target for the development of new antifungal drugs. In pursuit of this goal, we aimed to identify small molecules capable of inhibiting the DNA-binding ability of Zn2Cys6 TFs regulating PDR in yeast. An initial screen utilizing small molecule microarrays (SMM) was employed to identify compounds capable of binding the DNA-binding domain of Zn2Cys6 TF Pdr1p from Candida glabrata. In our initial SMM-based screen, a library of 15,000 different compounds was examined and yielded 76 unique compounds that specifically bound to Pdr1p. Small-molecule 'hit' compounds consisted of several different structural classes, including 44 azetidine and sulfonamide-based compounds. Future studies will examine the potential of these lead compounds for development of improved antifungal drugs in the treatment of IFIs.

“GeauxDock: an ultra-fast molecular docking package for computer-aided drug discovery”

Yun Ding
Louisiana State University
PosterID: 3

Computational modeling of binding drug to proteins has become an integral component of modern drug discovery pipelines. A typical application is structure-based virtual screening, which involves a large-scale modeling of pharmacological relevant associations between small molecules and their macromolecular targets. The desire to improve state-of-the-art motivated us to develop an ultra-fast ligand docking approach that uses Monte Carlo as the sampling method and features computations on modern supercomputers. Combined with an effective scoring function, this new method will provide accurate predictions at a high performance/cost ratio, which is a critical factor for large-scale virtual screening applications.

“A Provenance Management System for Tomography Data Processing and Visualization”

Jumao Yuan
Louisiana State University-Baton Rouge
PosterID: 4

In computerized tomography the algorithms used for data processing are constantly evolving and improving. Each research group has its own preferred platforms, be it Python, Mathematica, or MATLAB scripts, which can make it hard to integrate someone else's code into your data processing pipeline. These factors can also make it difficult to recall, years later, how any given dataset was processed. Which algorithm was used? Which version of software? Provenance management systems help to track the history of data processing. VisTrails, freeware developed at New York University, allows for the creation of modules that encompass each step of data processing and the creation of a history tree that tracks changes to the workflow as it is developed. Modules can call other programs, which can help integrate multiple pieces of software into a single workflow. The history tree allows for exploratory actions to be logged so that they can be easily returned to in the future. It is a great step towards making the development of tomography data processing more accessible and shareable. Here we will showcase two workflows we developed that show the utility of the VisTrails software. The first is an X-ray grating interferometry dataset of a foraminifera gathered at the Advanced Photon Source beam line. The second is a collection of multi-energy (12 keV to 32 keV) scans of a burnt flame retardant/polymer blend. A definite need was noted for an active sample position control system to account for the sample consumption and motion out of the field of view. The foraminifer data processing involves the evaluation of absorption, phase-contrast, and dark-field images from raw stepped-grating interferometry. The history tracking and visualization methods of VisTrails were useful in carrying out comparisons of various tomography reconstructions. It was also possible to integrate modules to export the dataset to a mobile visualization platform KiwiViewer (developed by KitWare), which allowed for facile viewing and sharing of the results. Data obtained for the burnt flame retardant presented different issues. The large and multiple data sets required extensive use of high performance computing---a system with solid-state storage and 12 core, 196 GB RAM node---and the management of files between local and remote servers. Also, while the reconstruction methods were similar to the foraminifera project, additional processing was needed in MATLAB and Mathematica to convert the absorptions near K-edge into relative volumen percent of constituent materials. Through the use of custom Python scripting modules, VisTrails was able to handle the switching between applications well. The modules that were developed for the processing of these two datasets should lead to the development of a custom package for our research group that will then be available to the VisTrails community. Easy sharing of packages allowed by VisTrails should encourage openness of data processing techniques between research groups.

“CEMAsuite: Degenerate Consensus Primer Design”

Courtney Lane
Louisiana State University, Department of Chemical Engineering
PosterID: 5

CEMAsuite, a consensus PCR primer design application first presented at the 2nd Annual LA Conference on Computational Biology & Bioinformatics, has been expanded to include degenerate (permutable) nucleotide handling and batch primer-template hybridization. Thus, granting the user total control over the degenerate primer construction process. Using a protein multiple sequence alignment, the application: generates a codon-equivalent multiple alignment (CEMA), scores each nucleotide position using one of four algorithms, and visualizes said scores. Next, a consensus primer set can be designed using Primer 3, an open-source primer design software capable of accounting for positional scores in its objective function, or by inspection. Score-weighted consensus primer design has proven to result in significantly more robust primers than unweighted design. The added features of CEMAsuite now allow the user to quickly estimate the stability, as a Gibbs free energy value, and view the primary structure of a primer template duplex for every CEMA DNA sequence. The hybridization algorithm uses traditional nearest-neighbor thermodynamics with adjustments for specific reaction conditions. A notification system highlights specific cases which are 'unlikely to amplify' based on two readily adjustable input thresholds. Default threshold values were obtained by subjecting PCR amplification results reported in literature to the hybridization algorithm. The degenerate nucleotide handling capability allows for the incorporation of selective degeneracy in the primers in order to obtain a more robust primer set while maintaining a low total degeneracy. The hybridization algorithm handles each primer permutation independently and presents descriptive statistics on degenerate hybridizations. The intent of this expansion is to aid in the design of a low-degeneracy primer set which is robust enough for the applicable assay, yet retains as much specificity as possible.

“Developing the DARE-REM science gateway for large-scale biomolecular conformational sampling with RESTMD”

Shayan Shams
Louisiana State University-Baton Rouge
PosterID: 6

We present our recent developmental effort for building the DARE-REM science gateway. The gateway aims to primarily offer a service of the scientific application, Replica Exchange Statistical Temperature Molecular Dynamics (RESTMD), along with other known replica exchange methods (REMs). RESTMD is a recently proposed novel method for enhanced sampling and has various advantages over other sampling methods such as REMD. This work also underscores our long term interest for establishing effective strategies to build versatile, extensible, lightweight, modular science gateways around the DARE framework. DARE- REM, therefore, addresses a core design goal of DARE-based gateways, the seamless use of heterogeneous distributed cyber-infrastructures (DCIs), which is essential for enabling extreme-scale computational tasks and data analytics. Among many novel contributions, the implementation of STMD is now available in two community molecular dynamics packages, CHARMM and LAMMPS, which is an attractive point for a wide range of life science researchers. The key workflow for replica exchange methods requires the execution of multiple replicas attempting a pairwise exchange along a simulation run, and is particularly implemented using Hadoop MapReduce. DARE-REM can submit a RESTMD or REMD job over various HPC cluster systems, Amazon EC2, our testbed CRON, and GENI. In addition to the capacity of utilizing these heterogeneous systems, with the CRON testbed system, interestingly, we demonstrate that RESTMD is capable of the scale-across scenario. Our experience with the GENI virtual laboratory, relatively unknown to the science gateway community, is also presented. The DARE-REM is accessible at http://dare.cct.lsu.edu, and will be further developed providing more user friendly features in addition to the core features described in this work.

“Use of docking method to model protein-protein interaction inhibition: Application to Mini-Proteins (cyclotides).”

Rushikesh Sable
University of Louisiana at Monroe
PosterID: 7

The aim of this project is to design and characterize stable, small CD2 protein derived peptide molecules with the grafted cyclotide technique. These grafted cyclotides are expected to inhibit CD2-CD58 protein-protein interaction which can prevent progression of autoimmune diseases. Cyclotide grafting strategy appends exceptional thermal and enzymatic stability to the molecule. Previously established cyclic peptide sequences used here to design the grafted sequences. We have used Sun flower trypsin inhibitor (SFTI) and Rhesus theta defesin (RTD) as a framework cyclotide structures for grafting. Five different grafted sequences designed using YASARA molecular modeling software. The specific aims of the project were: a) to evaluate the protein-protein interaction inhibition activity of cyclotides, b) to elucidate the stability profile of the cyclotides and c) to determine the rational model of grafted cyclotide binding activity by docking. For protein-protein interaction inhibition activity, a lymphocyte-epithelial cell adhesion assay was used. For the determination of serum stability SFTI 1-1 was incubated in human serum for 72 hours and for thermal stability assessment the change in elipticity (Circular Dichroism) data used after gradual increase in cyclotide temperature from 25 to 85 °C. AutoDock software was used to determine the virtual binding of these designed molecules to the CD58 and CD48 (protein present in rodents having 60% homology to CD58) protein structures by docking. Lowest docked energy structures were analyzed after scoring. Among five different grafted cyclotides, SFTI 1-1 showed promising cell adhesion inhibitory activity (IC50 ~ 100 nM). It also demonstrated good serum stability in human serum till 72 hrs, and was stable for high temperatures till 85 °C. Results of docking showed SFTI1-1 (grafted cyclotide) was efficiently binding to the CD58 and CD48 proteins with lower binding energies which inhibits the binding of CD2 to CD58/CD48 proteins.

“MOLECULAR DOCKING STUDIES OF DOXORUBICIN-PEPTIDOMIMETIC CONJUGATES TARGETING EXTRACELLULAR DOMAIN OF HER2 RECEPTOR”

Sandeep Palleria
LSU Ag center, Baton Rouge
PosterID: 8

Doxorubicin (DOX) is the cornerstone in the therapy of many carcinoma types. Unfortunately, the cytostatic effect of DOX in therapeutic doses is frequently insufficient; however, the use of higher DOX doses is limited by the development of systemic toxicity, especially cardiotoxicity. Many approaches have been used to deliver doxorubicin to the active site by conjugating the drug with peptides, antibodies, hormones, etc. In the current approach, we have attempted to conjugate doxorubicin to a peptidomimetic that is highly specific for HER2 positive cancer cells. The peptidomimetic we have designed (compound 5) is known to bind to extracellular domain of HER2 and inhibit protein-protein interaction of human epidermal growth factor receptors. Different strategies were used to synthesize the conjugate of compound 5 with doxorubicin. To optimize the linkage of doxorubicin to peptidomimetic we used computational docking method. Using the crystal structure of HER2 with 620 amino acid residues of the extracellular domain that represents domain IV of HER2, different conjugates with or without linkers were docked using AUTODOCK software to the region around domain IV. The lowest energy docked conjugate was synthesized in the lab using manual solid-phase synthesis method. The synthesized compound was purified by HPLC, analyzed by mass spectrometry and further biological assays were performed. Funding for this research was from the National Institute of General Medical Sciences of the National Institutes of Health under grant number 8P20GM103424.

“Real-Time PCR and High Resolution Melting Analysis (PCR-HRM) for Rapid Identification of Rifampin Resistance in Mycobacterium leprae Isolates and Clinical Samples from Brazil”

Sergio Araujo
Louisiana State University-Baton Rouge
PosterID: 9

Despite three decades of multidrug therapy (MDT) leprosy persists in many regions. Rifampin (RMP) is the crucial bactericidal agent in MDT. An increase in relapses and MDT treatment failures due to RMP resistance has been observed in Brazil, which together with the emergence of primary RMP-R Mycobacterium leprae, could undermine existing control measures. WHO recommends PCR-Direct DNA sequencing of RMP resistance determining region (RRDR) of the rpoB gene as a method for surveillance of RMP resistance in clinical isolates. However this technique is not available to most endemic communities and is not cost-effective for large sampling surveys. This project proposed to evaluate the Real-Time PCR high resolution melting analysis (PCR-HRM) as a screening tool for rapid identification of RMP resistance in clinical samples from Brazil, the second most highly endemic country for leprosy in the world. A panel of 14 reference strains, including those containing the most common RRDR mutations leading to RMP resistance, and RMP susceptible wild-type (WT) strains were utilized for optimization of the ML RMP PCR-HRM assay. DNA purified from 64 skin biopsies of leprosy patients who attended the CREDESH - National Reference Center for Sanitary Dermatology and Leprosy Clinic in Uberlandia, Brazil and tested positive for M. leprae DNA were included in this study. Results demonstrated that all characterized RMP-R strains were identified as distinct variants from the WT strain profile. DNA sequencing of the RRDR confirmed the RMP genotypes. Of the 64 clinical samples, we confirmed the WT genotype for 61 (95%) of these samples. Three samples (4.6%) contained the HRM variant profile. One of these (1.6%) had a confirmed RMP-R genotype. Taken together these results demonstrate the utility of PCR-HRM as a reliable screening tool for RMP susceptibility. The HRM is a closed-tube method which avoids post-PCR manipulation and can analyze up to 42 samples in a 96 well-based format within less than 3 hours. This enhances the applicability in endemic regions and reduces cost and time for RMP screening, thereby providing valuable information not only to improve patient treatment outcome but to the global context of leprosy drug resistance monitoring. We recommend that samples with HRM variant profiles be further evaluated by DNA sequencing.

“PsbP and PsbQ Interactions in Higher Plant Photosystem II”

Manjula Mummadisetti
Louisiana State University-Baton Rouge
PosterID: 10

Photosystem II (PSII) is a light-driven, water plastoquinone oxidoreductase. This important enzyme produces all of the oxygen on the planet, and is present in all oxygenic organisms. The oxygen-evolving complex (OEC), the site for water oxidation, is comprised of a Mn4CaO5 cluster. Extrinsic proteins (PsbO, PsbP and PsbQ) are present on the lumenal surface of PSII and stabilize the OEC by protecting the Mn-cluster from reduction and enhance PSII activity under physiological conditions. The PsbP and PsbQ proteins specifically maintain the calcium and chloride environment around the Mn-cluster to support optimal rates of oxygen evolution and are fundamentally important for plant PSII. In this study, protein cross-linking coupled to high-resolution mass spectrometry was used to study the protein structure and binding location of PsbP and PsbQ when bound to Photosystem II. We have identified that the N-terminus of PsbP, which is unresolved in the current crystal structure, interacts with the C-terminus of PsbP, forming a compact structure. We have also identified the binding location of PsbQ with respect to PsbP in photosystem II. Additionally crosslinked domains within PsbQ indicate that two copies of PsbQ are present, which associate in an antiparallel fashion. Our results place strong constraints on the structural organization of these proteins within the photosystem.
Reference: Mummadisetti, M.P., Frankel, L.K., Bellamy, H.D., Sallans, L., Goettert, J.S., Brylinski, M., Limbach, P.A., and Bricker, T.M. (2014). Use of protein cross-linking and radiolytic footprinting to elucidate PsbP and PsbQ interactions within higher plant Photosystem II. Proceedings of the National Academy of Sciences.

“A Graph-Based Approach to Targeted Drug Discovery”

Misagh Naderi
Louisiana State University-Baton Rouge
PosterID: 11

Due to extremely high costs of high-throughput screening, many drug discovery projects commonly employ inexpensive computations such as virtual screening to support experimental efforts. However, using libraries such as the ZINC database the vast majority of compounds will have a very low probability to exhibit the desired bioactivity for a given protein. Combinatorial chemistry methods can be used to produce a chemical universe of 1012 to 10180 drug-like compounds to be augmented to existing compound libraries. But, such chemical space is too large to be screened. Consequently, the trend in library design has shifted to produce a screening collection of compounds specifically tailored to modulate the function of a particular target or a protein family. Assuming that the organic compounds are composed of sets of rigid fragments, connected by flexible linkers, a molecule can be decomposed into its building blocks tracking their atomic connectivity. Our method synthesizes new compounds reconnecting these building blocks following the connectivity patterns via an exhaustive graph-based algorithm. In order to evaluate the performance, we have conducted a series of benchmarking calculations against the DUD-E dataset. First, in a self-benchmarking test, we validated the correctness of the algorithm. The object was to recover a molecule from its building blocks. Results show that our approach can efficiently rebuild more than 80% of the active molecules. 90% of the molecules were rebuilt in less than a minute, while half of them took only a fraction of a second to be reconstructed properly. The capability to discover novel scaffolds is also being assessed in a cross-benchmarking test. The procedure mimics a real application, where one expects to discover novel compounds based on a small set of already developed bioactive compounds. We are very optimistic that our effort can contribute to targeted drug discovery particularly in search for novel broad-spectrum antibiotics.

“A hypothesis-driven approach to identify functionally relevant transcription factor binding sites in ChIP-seq data”

Chris Gissendanner
University of Louisiana at Monroe
PosterID: 12

Chromatin Immunoprecipitation followed by high-throughput sequencing (ChIP-seq) is a frequently utilized approach to identify genomic transcription factor binding sites and target genes. We applied this technique to identify target genes of the NR4A nuclear receptor NHR-6 in the model organism C. elegans. NHR-6 regulates organogenesis through a DNA-binding mechanism and is necessary to promote cell cycle progression and cell differentiation during the organ formation process. Our ChIP-seq analysis identified 3,222 in vivo genomic binding sites for NHR-6. To identify key transcriptional programs regulated by NHR-6, we employed a multi-step, hypothesis-driven approach to filter binding site data. Our approach involved: 1) selecting binding sites most likely involved in cis regulation of gene expression; and 2) a two-step function-based filter utilizing database information from wormbase.org. The latter involved filtering based on predicted characteristics of NHR-6 target genes, including homology, expression and cellular function. Through this filtering protocol we established a set of 124 putative target genes hypothesized to be critical targets of NHR-6 during organogenesis. We tested the developmental functions of 82 of these putative target genes by RNAi and found that 63.4% exhibited similar loss of phenotypes as NHR-6. Interestingly, several of the genes identified function in cellular pathways that genetic interaction studies suggested were involved in the regulation of, or are regulated by, NHR-6. In addition, for most of the genes, we were able to get secondary confirmation that NHR-6 binds their regulatory regions. Utilization of MotifMutator, a novel motif discovery tool developed for this project, also revealed potential shared binding site motifs in the target gene promoters. Taken together, this integrated analysis of ChIP-seq data has revealed a mechanistic model for how NHR-6 regulates key cellular decisions during organogenesis.

“Genetic diversity and population structure of the Basmati rice (Oryza sativa L.) germplasm from North Western Himalayas using trait linked SSR markers”

Romesh Salgotra
Louisiana State University-Baton Rouge
PosterID: 13

Basmati, the unique aromatic quality rice, is a nature's gift to Indian sub-continent which has been grown in the foothills of the Himalayas for hundreds of years. Basmati rice is desirable in international market for its unique quality attributes but there is very little information available on genetic diversity of traditional basmati rice. Keeping the above in view, the present study was conducted to know the genetic base and population structure of basmati genotypes grown in NW Himalayas. The present study material consisted of 141 basmati rice accessions representing landraces, farmer's varieties, elite cultivars and breeding lines were using 40 highly polymorphic trait linked SSR markers. Total 112 alleles were detected by the aforesaid primers with a maximum and minimum frequency of 5 and 2, respectively. The maximum and minimum PIC values were found to be 0.63 and 0.17 for the primers RM206 and RM213, respectively, with an average PIC value of 0.41/marker. The average genetic similarity coefficient from all possible combinations was found to be 0.60. Phylogenetic-based cluster analysis of the SSR data, based on distance, divided all genotypes into four groups (I, II, III and IV), whereas model-based clustering method divided these genotypes into five groups (A, B, C, D and E). The basmati rice germplasm were found to have moderate diverse genetic base and high population structure. The diverse genotypes and highly polymorphic trait linked SSR markers identified in the present study will be utilized for the identification of QTLs/genes for different economically important traits and molecular breeding programme of rice in future.

“Trends in Demographic and Drug Screening Results among LSUHSC-Shreveport Emergency Department Patients during 1998-2011”

Phillip Kilgore
Louisiana State University-Shreveport
PosterID: 14

Considerable difficulty is encountered when evaluating drug-usage trends in a metropolitan setting; some methods utilize self-reporting, while longitudinal studies are often restricted to small cohorts. We instead obtained de-identified urological drug screening data from LSUHSC-Shreveport that covers a 13-year period (1998-2011) and consists of 111,359 visits by 71,311 patients. Though tests for specific drugs varied due to periodic changes in drug testing protocol, these tests consistently tested for amphetamines, barbiturates, benzodiazepines, cannabinoids, cocaine, and opiates. We constructed demographic profiles based on patient's ethnicity, gender, and age classification to determine detection patterns across each class. The majority of visits to LSUHSC-Shreveport include the African American (AA) and Caucasian (Cc) samples; AA represented the plurality of patients. Most notably, we observed a striking increase of positive tests for cannabinoids in the 0-4 age group; ~10-fold and ~5-fold increases were observed for this age group for AA and Cc respectively. Cocaine positive tests were predominantly found in AA children and adults; however, cocaine positives decreased during this period. Tests for both benzodiazepines and opiates were higher in Cc regardless of age; although benzodiazepine positives decreased in the general population over the past 10 years, this number had increased ~3-fold for both AA and Cc in the 0-4 age group, while opiate positives increased ~3-fold over the 13 year period (with greater positives in Cc). We were also able to use this data to construct a drug history of some patients, finding that 7,721 patients had a positive test on at least two separate visits. Of these patients, cannabinoids, cocaine, and benzodiazepines were most likely to yield a positive result (25.5%, 17.6%, and 16.2% respectively). Work is ongoing to determine this data's applicability respecting the evaluation of the gateway hypothesis.

“Can Machine Learning Increase the Efficiency of Interpreting Radiology Reports?”

Kenneth Smith
Louisiana State University-Shreveport
PosterID: 15

Machine learning techniques applied to clinical informatics to help with the diagnosis, treatment, and prognosis of a number of diseases could alter a doctor's path to diagnosis and effectively increase the outcomes of patient care. We analyzed text radiology reports and grouped them into both abnormal and normal sets. An analysis of one hundred thousand reports was done using a training set of approximately twenty thousand reports to train a Naïve Bayesian classifier to sort them. Since an initial prior assumption of the relationship between abnormal and normal sets was unknown the first part of the study was an analysis of the data to estimate the prior assumption. This was done by using statistical analysis to find the rarest words in the training set and using a percentage of the rare words as a guideline to estimate that a report was abnormal. The next step was the main analysis of the data using the Bayesian classifier and accuracy testing. The level of accuracy obtained was very high, especially considering that the prior assumption was unknown.