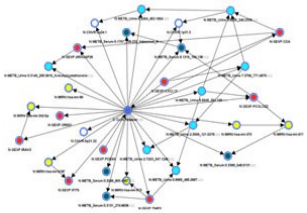
Showcase of Medical Discoveries

Biomedical Informatics Research



Wednesday, September 21, 2016
4:30—6:00 p.m.

A Wine & Cheese Reception of UAMS Investigators Discussing their Research and Discoveries.

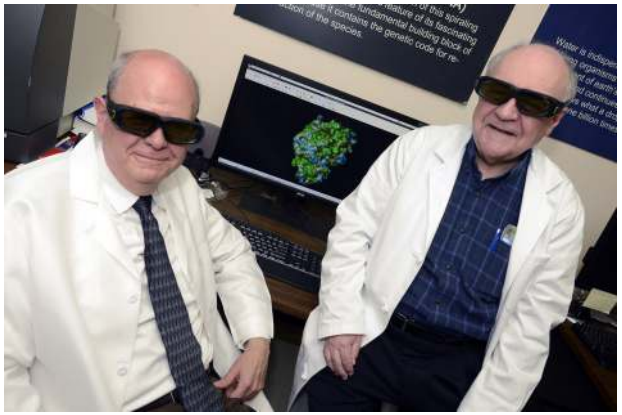
Winthrop P. Rockefeller Cancer Institute
10th Floor Rotunda

UAMS
COLLEGE OF MEDICINE
UNIVERSITY OF ARKANSAS FOR MEDICAL SCIENCES

UAMS Office of Research

Poster #1

Applications of Molecular Dynamics to the Design of Radioprotectant Tocotrienols with Enhanced Bioavailability



Cesar M. Compadre, Awantika Singh, Shraddha Thakkar, Guangrong Zheng, Philip J. Breen, Sanchita Ghosh, Mahmoud Kiaei, Marjan Boerma, Kottayil I. Varughese and Martin Hauer-Jensen

There is a pressing need to develop safe and effective radioprotector/radiomitigator agents for use in accidental or terroristic radiological emergencies. Naturally occurring vitamin E family constituents, termed tocols, include members, such as the tocotrienols. The tocotrienols have remarkable radiation-protection properties with minimal toxicity. Unfortunately the tocotrienols have very short plasma half-lives and require dosing at very high levels to achieve necessary therapeutic benefits. The short elimination half-life of the tocotrienols is related to their low affinity for the α-tocopherol transfer protein (ATTP), the protein responsible for maintaining the plasma level of the tocols. In an effort to uncover tocotrienol analogues that would have enhanced binding to ATTP, we developed a computational approach, for the virtual screening of those analogues. The approach we developed is based in molecular dynamics simulations and overcomes the limitations of other computational approaches. Based on this approach, we developed of a new class of vitamin E analogues, the tocoflexols, which maintain the superior bioactivity of the tocotrienols with the potential to achieve the longer half-life and larger AUC of the tocopherols.

Poster #14

What is Life? A Set of 500 Conserved Functional Domains

Visanu Wanchai, Arvind Ramanathan, Intawat Nookaew, David Foutch, Preecha Pratumcharoenpol, Se-Ran Jun, Trudy Wassenaar, Mike Leuze, and David W. Ussery

In 1995, the first two bacterial genomes were sequenced, and comparison of those two genomes resulted in a set of about 250 conserved proteins, initially deemed as the 'minimal set of proteins for all of life'. As more genomes were sequenced, the number of conserved proteins dropped, from about a hundred proteins in 20 genomes, to a mere 31 proteins found conserved in about 200 genomes, to zero proteins conserved across a thousand bacterial genomes. This is an example of a 'good failure' in that the unanticipated results points to a need for more careful analysis of whether the question of what does 'conserved protein' mean?

We have rephrased the question to look for conserved functional domains (using PfamA domains) across more than 70,000 bacterial genome sequences, and find a set of about 500 core PfamA domains that are conserved in at least 99% of the genomes. These conserved domains include all of the basic functions for life, such as polymerases for DNA and RNA synthesis, protein synthesis (all 54 ribosomal proteins and rRNA genes) and basic metabolic processes. We compare the proteins with these functional domains with sets of 'essential genes' determined from transposon mutagenesis experiments, as well as the basic set of functions predicted from 'first principles'.

Poster #13

Monitoring a Changing Profession: Clinical Research Data Management

Meredith Zozus

We undertook a survey to assess and refine competencies for the clinical research data management profession. Based on prior work developing and maintaining a practice standard and processional certification exam, a survey was administered to a captive group of Clinical Research Data Managers to assess professional competencies, types of data managed, types of studies supported and necessary foundational knowledge. Respondents confirmed a set of 91 professional competencies. As expected, differences were seen in job tasks between early-to-mid career and mid-to-late career practitioners. Respondents indicated growing variability in types of studies for which they managed data and types of data managed. Respondents adapted favorably to the separate articulation of professional competencies versus foundational knowledge. The increase in the types of data managed and variety of research settings in which the data are managed indicate the need for formal education in principles and methods that can be applied to different research contexts, i.e., formal degree programs supporting the profession, and stronger links with the Informatics scientific discipline, and Clinical Research Informatics in particular.

The results document the scope of the profession and will serve as a foundation for the next revision of the Certified Clinical Data Manager (CCDMTM) exam. A clear articulation of professional competencies and necessary foundational knowledge may inform content of graduate degree programs or tracks in areas such as Clinical Research Informatics that will develop the current and future Clinical Research Data Management workforce.

Poster #2

Comparative Genomic Evidence of Zika Virus Epidemiology

.

Se-Ran Jun, Intawat Nookaew, David Ussery

The Zika virus (ZIKV) is an emerging human-pathogenic flavivirus transmitted by mosquitoes, which has shown a similar spread pattern to Dengue and Chikungunya. For genomic understanding of Zika virus epidemiology, we performed phylogenetic comparative analysis and adaptive evolution analysis of about a hundred Zika virus genomes in comparison with other flavivirus genomes. The maximum likelihood tree of flavivirus complete genomes shows clear distinction between species even though the Zika virus has shown antibody cross reactivity and similar clinical presentation with Dengue and Chikungunya. The genome tree of Zika virus rooted by root-to-tip regression analysis displayed that the Zika virus identified in French Polynesia, 2013 is an origin of China outbreaks, 2016 showing better agreement with the timeline of outbreaks than the tree rooted to Spondweni virus. Phylogenetic comparison of genes and whole genome implied that NS5 might be of indicative of molecular evolution of whole genome. We identified sites in genes C, NS1, NS3, NS4B, and NS5 subject to positive selection, and sites under positive selection along the lineage of outbreaks between 2007 and 2016.

Poster #3

Empowering Information-Driven Healthcare

William Golden, Tyler Vodehnal & Marlo Harris

Arkansas became the first state to successfully enact a statewide payment reform initiative through the implementation of episode-based payments. By incentivizing providers to deliver care via a cost-effective system, Arkansas Improved healthcare quality and decreased Medicaid cost growth. In 2014, Arkansas Medicaid implemented the Patient-Centered Medical Home (PCMH).

The PCMH program assists primary care providers (PCPs) (least 300 beneficiaries) in their transition to a PCMH while rewarding high-quality, coordinated care. PCMH practices (least 5,000 beneficiaries) with approved activities, metrics and costs below statewide thresholds are eligible to receive shared savings up to 50 percent of the total shared savings.

To help providers manage their PCMH program, we developed an advanced data analytic tool that provides progress of beneficiary care, metrics and the costs associated with healthcare services. The analytics tool uses the claims data and beneficiary risk scores to report the risk-adjusted total cost of care:

- Metric target determination: Based on statistical analysis of statewide and national data
- Quarterly PCMH reports: Historical and current performance against statewide cost thresholds, benchmarks and metric targets
- Dynamic dashboard: Drill-down capabilities for individual PCMH performance and trending analysis
- Beneficiary and claim-level reports: Assist PCMH customer service center

Poster #12

Scientific and Methodological Advancements in Liquid Biopsies to Further the Development of Lung Cancer-based Precision Medicine

Donald J. Johann, Horacio Gomez-Acevedo, Erich Peterson, Jason Liem, Meei Liu, Kostas Arnaoutakis, Matt Steliga, Ikjae Shin, Dong Yoon, and Laura Hutchins

Next generation sequencing (NGS) is rapidly changing the manner in which biomedical research is performed and clinical medicine is practiced. NGS-based clinical assays are now providing a new and mechanistic manner of viewing human disease. Liquid biopsies have the potential to change medical oncology practice by providing a rapid and convenient method to interrogate DNA mutations present in the cell free DNA (cfDNA) fraction of blood. This tremendously reduces associated risk to patients versus most standard cancer-based biopsies. It also allows for more frequent reassessment over time, especially for monitoring of therapeutic response, or signs of resistance.

Lung cancer-based precision medicine is the focus in our advanced clinical trial, which utilizes a co-clinical trial methodology. The co-clinical trial platform is an innovative approach to improve the design, speed and outcomes of rational and personalized cancer treatments. ***Key objectives of our science and clinical trial include***: **i)** elucidation of genomic background metrics, **ii)** sensitive and specific identification of cell free DNA mutations, **iii)** further refining the liquid biopsy methodology along with the incorporation of a co-clinical trial approach, and **iv)** the development of novel and sophisticated bioinformatic tools for the processing, analysis, and integration of this large and complex "big data."

Poster #11

Microbiomics: Linking Human Gut Microbiome with Human Health and Disease

Intawat Nookaew, Visanu Wanchai, Preecha Pratumcharoenpol, Se-Ran Jun, and David Ussery

Gut microbiota is altered in different individuals depending on the inherited bacteria and composition of diet intake. There have been many studies recently linking changes in the microbiome composition with diseases and immune response in humans. Changes of gut micro-biota alter fermentation processes and the energy harvest from the gut ecosystems, leading to modulation of host-microbiome metabolic interactions. These will affect the host homeostasis and can contribute to disease development or to a more healthy state. With high throughput sequencing technologies, it is possible to generate deep sequencing of metagenome samples that allow the identification of community structures and functional profiles of individuals. However, data acquisition and analysis is challenging. We present our developed tools such as Functional And Taxonomic analysis of Metagenomes (FANTOM), Metagenomic Data Utilization and Analysis (MEDUSA), Metagenomic Species (MGS) through co-abundance gene groups (CAGs). These tools enable us to gain insight into the role of gut microbiota to the human body. We also present the association of gut microbiota on type2 diabetes (T2D) and variation of genetic element in complex human gut metagenome samples through our computational tools.

Poster #4

SemanticSurvey - A Questionnaire Tool to Collect Semantically-rich Medical Data

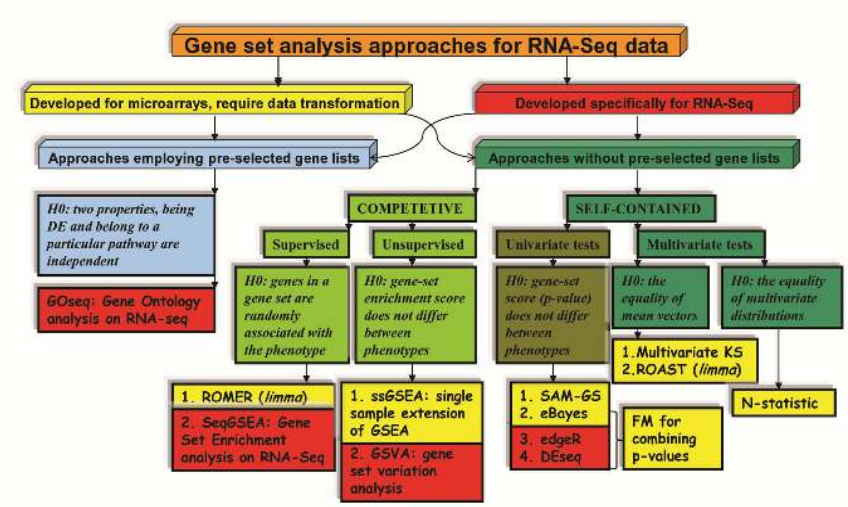
Joseph Utecht, John Judkins, Mathias Brochhausen

Collecting medically-relevant information using questionnaires is ubiquitous in both clinical practice and research. Many of these questionnaires are created using survey applications, such as Lime Survey or Survey Monkey. However, the biggest problem is how to retrieve data collected via questionnaire against the often evolving and shifting research questions pursued by a research project or needed for clinical purposes. Storing data in relational databases often restricts querying the data based on the underlying data schema and if often not very flexible if new attributes or values need to be added.

Semantic Web Technologies provide the means to store data in a schema -free manner. New attributes and new entities can be added anytime. Querying the data is not restricted by a data schema and relies on relationship between entities (e.g. a patient ID and a lab test result). We have developed a generic survey application that allows storing information based on answers provided by users in RDF, a semantic web standard. This tool allows the development of questionnaires in all domains of biomedicine, including clinical research. This tool was developed in the NIGMS-funded CAFÉ project (R01GM111324).

Poster #5

Performance Evaluation for Gene Set Analysis Approaches for RNA-seq data

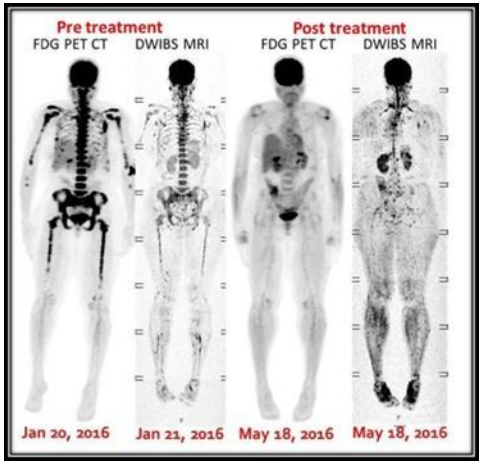


Yasir Rahmatallah and Galina Glazko

Transcriptome sequencing (RNA-seq) is steadily replacing microarrays for high-throughput studies of gene expression. Gaining insights into the biological processes underlying phenotypic differences is the major challenge in analyzing the expression of thousands of genes. For this purpose, gene set analysis (GSA) emerged as the method of choice because it incorporates pre-existing biological knowledge in functionally related gene sets into the analyses. We consider several GSA approaches that were adapted from microarrays practice or specifically designed for RNA-seq data. We perform a comprehensive evaluation of their performance in terms of error rate, power, robustness to sample size/ heterogeneity, and sensitivity to selection biases on simulated and real data. We show that the performance of various methods depends on the statistical hypothesis they interrogate and not on the platform type for which they were designed for. We found that competitive methods have lower power as well as robustness to samples' heterogeneity than self-contained methods, leading to poor results reproducibility. Also the power of unsupervised competitive methods is highly affected by the balance between up- and down-regulated genes in tested gene sets. Our evaluation provides a concise guideline for selecting proper GSA approaches, and warns against using others in their current form.

Poster #10

Comparison of the Accuracy of Whole Body FDG PET CT and Whole Body Diffusion Weighted MRI Imaging with Background Subtraction (DWIBS) in Multiple Myeloma.



Kristen Schemel, Rohan Samant, Leo Rasche, Amy Buros,
Faith Davies, and Thomas Eluvathingal

Introduction: Studies have shown that FDG PET CT and DWIBS are valuable in evaluating bone and bone marrow (BM) abnormalities in multiple myeloma, but literature is lacking in a direct comparison of these two techniques. The purpose of this study is to compare the accuracy of FDG-PET CT and DWIBS in multiple myeloma.

Materials and Methods: Retrospective review of a sample of 86 [45M, 65 yr. (37-85)] patients; initial evaluation of suspected myeloma, monoclonal gammopathy of uncertain significance (MGUS), smoldering myeloma (SMM) and treated patients in complete remission (CR). All patients underwent a FDG-PET CT and DWIBS within 3 days of each other and a bone marrow biopsy within a week. Data collected from both FDG -PET CT and DWIBS MRI imaging included the number of focal lesions (<5, 5-20, >20), as well as diffuse involvement of bone marrow.

Results & Conclusion: Positive imaging evidence for myeloma was defined as a BM/liver FDG uptake ratio of greater than 1.2 (or signal greater than spleen on DWIBS) and/or presence of one or more focal lesions. Both PET (91%, 94%) and DWIBS (87%, 92%) had very high specificity and positive predictive values. More important is the fact that, PET CT (97%) and DWIBS (91%) correctly identified all patients without active myeloma [CR (6/86), MGUS (16/86), and SMM (10/86)].

Poster #9

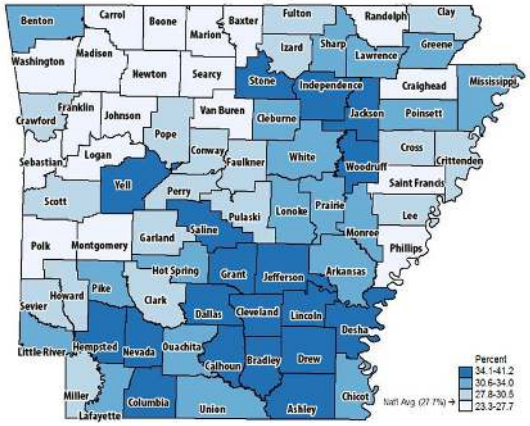
Data Standardization for Regulatory Decision-Making: Development of Data Standards in Mental Health, Cardiology, and Infectious Diseases

Anita Walden, Meredith Zozus, PhD

Federal requirements in the United States mandate sharing of research data, meaningful use of health information technology, and data standardization for regulatory review of marketed therapeutics. These requirements are predicated on the assumption that both healthcare organizations and the public will benefit from the enhanced secondary use of healthcare data. Because necessary standards are lacking across most clinical therapeutic areas, large-scale efforts are underway to create authoritative, consensus-based, and publically available standard data element sets. We report development and demonstration of methodology for national and international Data Element standardization within Health-Level Seven (HL7) in Mental Health, Cardiology and Infectious Diseases. Thus far, the work has resulted in nine and counting national and international data standards

Poster #6

Geographic Variation in Diagnosed Chronic Medical Conditions in the Arkansas Medicare Population.



T. Mac Bird, PhD, Kenley Money, MS

Background**:** As many as 8 out of 10 Medicare beneficiaries have at least one chronic medical condition. Disease burden is not evenly distributed geographically, with considerable variation existing county by county.

Methods**:** Data for this study came from the Arkansas All-Payer Claims Database (APCD). The APCD is the legislatively mandated central repository for health insurance claims data from all major public and private health insurance providers in Arkansas. For this study, only Medicare beneficiaries were included. Specific algorithms used to identify the chronic conditions in this study came from the Medicare Chronic Conditions Data Warehouse. Selected conditions include coronary heart disease, hypertension, stroke, depression, chronic obstructive pulmonary disease, and breast cancer.

Results**:** Coronary heart disease, hypertension, and stroke were most prevalent in eastern and southern Arkansas. Depression was most prevalent in the northeast, central, and northwest regions of the state. COPD was most prevalent in the northeast and southwest regions. Breast cancer was most prevalent in central Arkansas.

Conclusions**:** While Arkansas ranks higher than the national average in the prevalence of many chronic medical conditions, there is considerable variation at the county level. Understanding the geographic patterns of disease burden can aid policy makers in crafting more efficient, targeted policies and public health interventions.

Poster #7

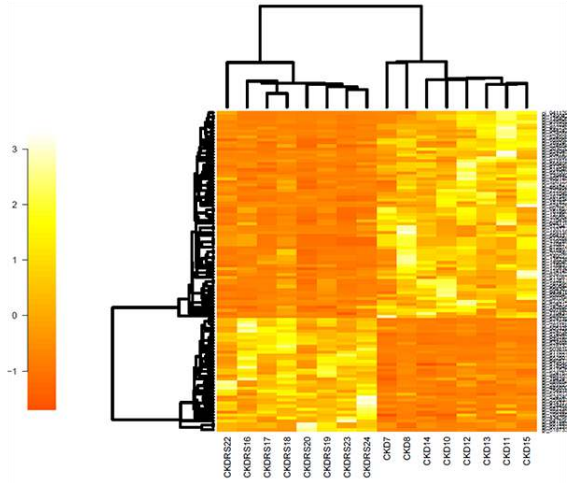
Integrative Analysis of Microbiome and Metabolome of the Porcine Gut Reveals Diet-associated Interrelationships

Sudeepa Bhattacharyya, Brian Piccolo, Kelly Mercer, Sree V Chintapalli, Kartik Shankar, and Laxmi Yeruva

Integration of multiomic data provides a significant capability of capturing the "big picture" in investigating the underlying biological mechanisms and their associations with phenotypes and disease. In this study, we used 16S ribosomal RNA amplicon sequencing and untargeted metabolomics to characterize the small intestinal microbiome and metabolomic repertoire of neonatal piglets fed either porcine breast milk (n = 6), milk-based infant formula (n = 11) or soy-based infant formula (n = 11). Signed weighted metabolite co-abundance correlation networks were constructed in which metabolites that were similar in their profiles were clustered into modules and summarized by the "module eigenvector" (which is, the first principal component of the metabolite abundances across the subjects). The module eigenvectors were used to correlate to the microbial taxonomic abundances in order to examine the relationship between the gut microbiome and metabolomic changes associated with diet groups. Our analysis identified metabolic signatures in the GI tract of pigs fed breast milk versus formula. Importantly, these changes were highly associated with changes in gut microbial diversity.

Poster #8

Metaproteomics Data Processing Pipe-Line: Application to the Study of Function of Gut Microbiome in Chronic Kidney Disease



Zybailov, B., Rahmatallah' Y., Glazko, G., Byrum S., Orr L., Tackett, A., Mackintosh, S., Edmondson, R., Keiffer, D., Adams, S., Viziri, N., Arthur, J.

Chronic kidney disease (CKD) is a growing health problem: 13% of adults in the US have CKD. CKD increases urea levels in bodily fluids leading to a dominance of urease-containing bacteria in the gut. Such dysbiosis results in decreased consumption of nitrogenous waste and erosion of the epithelial barrier resulting in bacterial toxins promoting inflammation. Our project is aimed at mapping the gut microbiome-host interactions, identifying microbiome-derived prognostic biomarkers of CKD progression, and at developing methods to target the gut microbiome for CKD therapies and prophylaxis. Resistant starch (RS) is a type of pre-biotic that is not fully broken down and absorbed, but rather turned into short-chain fatty acids by intestinal bacteria. In the microbiome analysis of adenine-induced rat model of CKD, it has been demonstrated, that RS diet reduces plasma toxins, and decreases gut dysbiosis. Our new quantitative metaproteomics data processing pipe-line catalogued RS-induced microbiome changes at unprecedented resolution. More than 400 assembled proteins were identified as differentially abundant. Proteins significantly up-regulated in RS-fed rats were mostly of bacterial origin, while proteins, down-regulated in RS-fed rats were mostly host proteins. Our future plans include validating our findings using germ-free mice, and launching Phase-II human clinical trials of the resistant starch diet in advanced CKD patients.