The 9th Annual Conference of the Arkansas Bioinformatics Consortium (AR-BIC)

AR-BIC 2023

We are an Arkansas Collaborative Community in Bioinformatics Research

https://ar-bic.aralliance.org/

Biointormatics, Big Data, AI, and Public Health: An Integrated World

March 13-14, 2023

Conference Center
Wyndham Riverfront Little Rock
North Little Rock, Arkansas

Conference Sponsors

ArkansasResearchAlliance
Coming together to move Arkansas forward.

FDA U.S. FOOD & DRUG ADMINISTRATION
NATIONAL CENTER FOR TOXICOLOGICAL RESEARCH

UAMS
UNIVERSITY OF ARKANSAS FOR MEDICAL SCIENCES

UALR
UNIVERSITY OF ARKANSAS AT LITTLE ROCK

ARKANSAS INC

ASU
ARKANSAS STATE UNIVERSITY

UNIVERSITY OF ARKANSAS

Arkansas Biosciences Institute

UNIVERSITY of ARKANSAS AT PINE BLUFF
1873

AR-BIC 2023: Bioinformatics, Big Data, AI, and Public Health: An Integrated World

Established in 2014, the Arkansas Bioinformatics Consortium (AR-BIC) is an Arkansas-centric community that facilitates communication and collaboration among researchers to leverage the state's expertise and resources in data sciences and bioinformatics.

AR-BIC is hosting its 9th annual conference this spring. It focuses on the integration of bioinformatics, big data, and artificial intelligence (AI) as a system to improve public health and advance medical care. Today's healthcare relies on big data streams (e.g., image, graphic, and omics data) from emerging technologies. These types of data necessitate integrative strategies because they tend to be complex, multi-dimensional, and result from efforts to automate digitization of legacy reference studies. Meanwhile, AI approaches to synthesize, interpret, and leverage data have demonstrated significant impact across a broad range of scientific disciplines. Data connectivity, computational resources, and new/advanced bioinformatic strategies fuel the rise of AI, providing new insight into underlying mechanisms of human health and disease and their susceptibility to exogenous perturbations. Thus, this meeting will provide a platform to present and discuss the current state-of-the-art practice and on-going efforts in applying AI in healthcare and enabling efficient data mining to promote public health. This 1.5-day conference will be held at Wyndham Riverfront Little Rock, March 13-14, 2023, https://ar-bic.aralliance.org/.

Organized and supported by:

Arkansas Bioinformatics Consortium (AR-BIC)

Conference sponsors and acknowledgements:

* Arkansas Biosciences Institute (ABI)

* Arkansas Economic Development Commission (AEDC)

* Arkansas Research Alliance (ARA)

* Arkansas State University (ASU)

* Food and Drug Administration (FDA)

*University of Arkansas (UA)

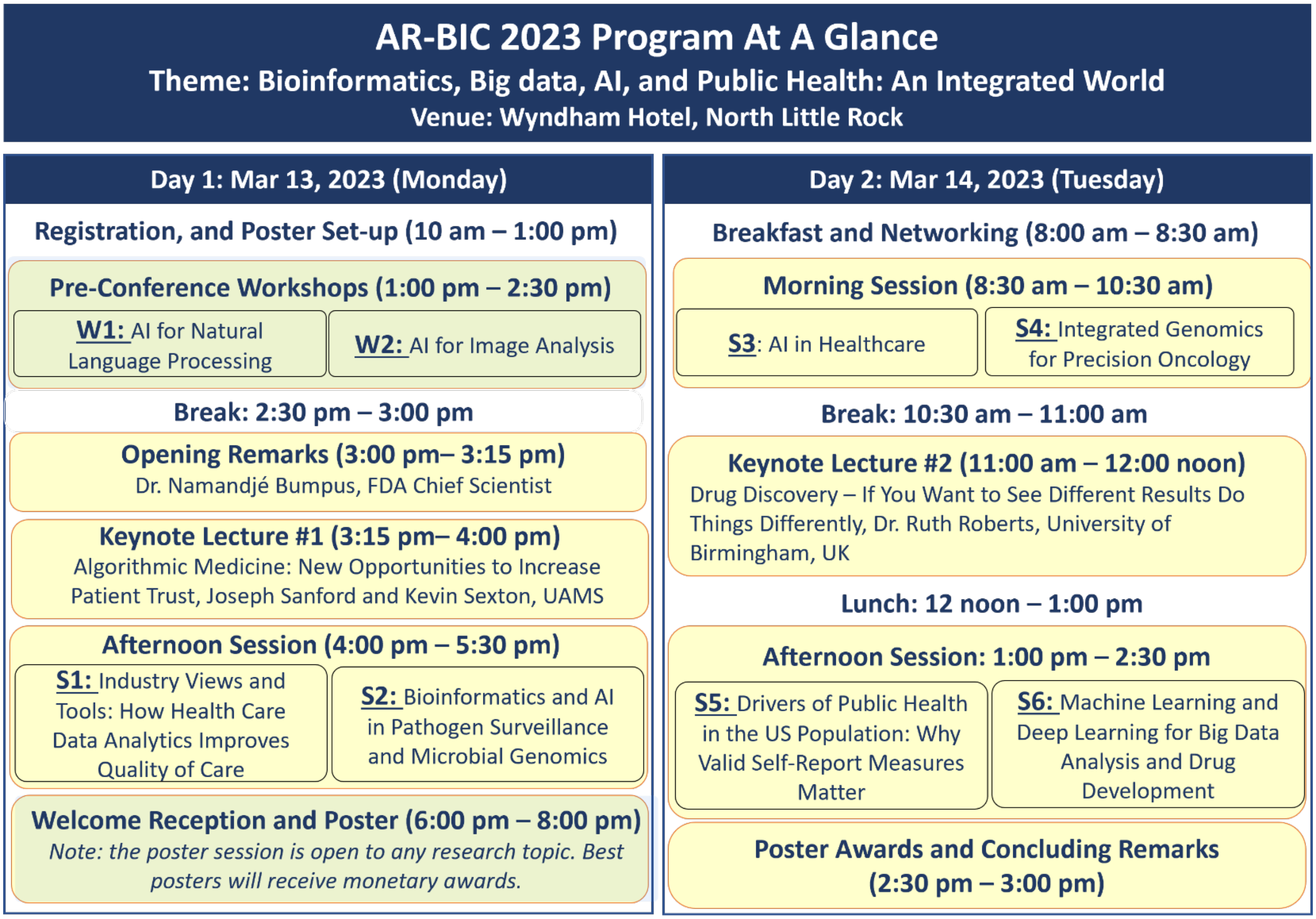
*University of Arkansas at Little Rock (UALR)

* University of AR for Medical Sciences (UAMS)

* University of Arkansas at Pine Bluff (UAPB)

**Funding for this conference was made possible, in part, by the Food and Drug Administration through grant 1R13FD005304-02, views expressed in written conference materials or publications and by speakers and moderators do not necessarily reflect the official policies of the Department of Health and Human Services; nor does any mention of trade names, commercial practices, or organization imply endorsement by the United States government.*

TABLE OF CONTENTS



AR-BIC 2023 Program At A Glance

Theme: Bioinformatics, Big data, AI, and Public Health: An Integrated World

Venue: Wyndham Hotel, North Little Rock

Day 1: Mar 13, 2023 (Monday)

Registration, and Poster Set-up (10 am – 1:00 pm)

Pre-Conference Workshops (1:00 pm – 2:30 pm)

- **W1:** AI for Natural Language Processing
- **W2:** AI for Image Analysis

Break: 2:30 pm – 3:00 pm

Opening Remarks (3:00 pm– 3:15 pm)
Dr. Namandjé Bumpus, FDA Chief Scientist

Keynote Lecture #1 (3:15 pm– 4:00 pm)
Algorithmic Medicine: New Opportunities to Increase Patient Trust, Joseph Sanford and Kevin Sexton, UAMS

Afternoon Session (4:00 pm – 5:30 pm)

- **S1:** Industry Views and Tools: How Health Care Data Analytics Improves Quality of Care
- **S2:** Bioinformatics and AI in Pathogen Surveillance and Microbial Genomics

Welcome Reception and Poster (6:00 pm – 8:00 pm)
Note: the poster session is open to any research topic. Best posters will receive monetary awards.

Day 2: Mar 14, 2023 (Tuesday)

Breakfast and Networking (8:00 am – 8:30 am)

Morning Session (8:30 am – 10:30 am)

- **S3**: AI in Healthcare
- **S4:** Integrated Genomics for Precision Oncology

Break: 10:30 am – 11:00 am

Keynote Lecture #2 (11:00 am – 12:00 noon)
Drug Discovery – If You Want to See Different Results Do Things Differently, Dr. Ruth Roberts, University of Birmingham, UK

Lunch: 12 noon – 1:00 pm

Afternoon Session: 1:00 pm – 2:30 pm

- **S5:** Drivers of Public Health in the US Population: Why Valid Self-Report Measures Matter
- **S6:** Machine Learning and Deep Learning for Big Data Analysis and Drug Development

Poster Awards and Concluding Remarks (2:30 pm – 3:00 pm)

GENERAL INFORMATION

Venue and Date:

- Wyndham Riverfront Hotel in North Little Rock: 2 Riverfront Pl, North Little Rock, AR 72114
- March 13-14, 2023 (Monday and Tuesday)

AR-BIC Advisory Council 2022-2023

- Arkansas Biosciences Institute (ABI): **Bobby McGehee**
- Arkansas Economy Development Commission (AEDC): **Jennifer Fowler**
- Arkansas Research Alliance (ARA): **Bryan Barnhouse, Julie LaRue, Douglas Hutchings, Amy Hopper Swan**
- Arkansas State University (ASU): **Thomas Risch**
- Center for Drug Evaluation and Research, Food and Drug Administration, **Shraddha Thakkar**
- National Center for Toxicological Research (NCTR): **Tucker Patterson, Weida Tong, Dongying Li**
- University of Arkansas (UA): **Margaret McCabe**
- University of Arkansas at Little Rock (UALR): **Brian Berry**
- University of Arkansas for Medical Sciences (UAMS): **Shuk-Mei Ho**
- University of Arkansas at Pine Bluff (UAPB): **Mansour Mortizavi**

Scientific Program Committee (2022-2023):

- Arkansas Research Alliance (ARA): **Bryan Barnhouse, Douglas Hutchings, Julie LaRue, Amy Hopper Swan**
- Arkansas State University (ASU): **Asela Wijeratne**
- Center for Drug Evaluation and Research, Food and Drug Administration, **Shraddha Thakkar**
- National Center for Toxicological Research (NCTR): **Steve Foley, Huixiao Hong, Dongying Li, Weida Tong (Chair), Joshua Xu**
- University of Arkansas (UA): **Douglas Rhoads and Samantha Robinson**
- University of Arkansas at Little Rock (UALR): **Phil Williams**
- University of Arkansas for Medical Sciences (UAMS): **Keith Bush, Shuk-Mei Ho, David Ussery,**
- University of Arkansas at Pine Bluff (UAPB): **Mansour Mortazavi and Grace Ramena**

Point of Contact:

- Logistics: Dongying Li (Dongying.Li@fda.hhs.gov) and Julie LaRue (jlarue@aralliance.org)
- Scientific Program: Weida Tong (Weida.Tong@fda.hhs.gov)

PROGRAM AGENDA

DAY 1 – MARCH 13, 2023

10:00AM – 1:00PM **Registration and Poster Setup**

1:00PM – 2:30PM **Pre-Conference Workshops (Parallel)**

- **Workshop 1: AI for Natural Language Processing** (Dr. Xiaowei Xu, Professor, Department of Information Science, University of Arkansas at Little Rock)
- **Workshop 2: Deep Learning Based Analysis of Histopathological Images of Breast Cancer** (Dr. Joe Zhang, Professor, School of Computing Sciences and Computer Engineering, University of Southern Mississippi)

3:00PM – 4:00PM **Opening Ceremony**

- **Opening Remarks:** Dr. Namandjé Bumpus, FDA Chief Scientist
- **Keynote Lecture**: **Algorithmic Medicine: New Opportunities to Increase Patient Trust** Drs. Joseph Sanford and Kevin Sexton, University of Arkansas for Medical Sciences

4:00PM – 5:30PM ***Session 1 and Session 2 (Parallel)***

SESSION 1: INDUSTRY VIEWS AND TOOLS: HOW HEALTH CARE DATA ANALYTICS IMPROVES QUALITY OF CARE

Chair: Bryan Barnhouse, President & CEO, Arkansas Research Alliance

Session Overview:

Organizations across Arkansas use data analytics and computing to increase revenue, reduce costs, create operational efficiencies, and improve overall performance. Data analytics used by hospitals, clinics, and other health care-affiliated organizations do all of that and save lives. AR-BIC attendees interested in learning how some of Arkansas' leading companies leverage the power of health care data analytics should attend this session. The speakers will review the challenges, opportunities, tools, and applications to convert large, accurate data into actionable and innovative solutions in their quest to improve the quality and efficiency of care and services for better patient and client outcomes.

4:00 – 4:30PM ***Birds of a Feather: Health Events Lead to Changes in Household Healthcare Utilization,*** Drs. Aaron M. Novotny and Elizabeth Parker, Arkansas Blue Cross and Blue Shield

4:30 – 5:00PM ***Using the Arkansas Healthcare Transparency Initiative (HTI) to Estimate the Cost of Smoking with IHME's Population-Attributable Fractions,*** Dr. Nichole Stanley and Kenley Money, Arkansas Center for Health Improvement

5:00 – 5:30PM ***Reduction of Central-Line Associated Bloodstream Infections, with Less Staff and More Patients,*** Dr. Amanda Novack, Baptist Health Center

SESSION 2: INDUSTRY VIEWS AND TOOLS: HOW HEALTH CARE DATA ANALYTICS IMPROVES QUALITY OF CARE

Co-Chairs: Dr. Douglas Rhoads, University of Arkansas; Dr. Steven Foley, National Center for Toxicological Research, US FDA

Session Overview:

We will explore different aspects of bioinformatics and the possible application of AI for surveillance, prevention, and treatment for microbial pathogens.

4:00 – 4:30PM ***Using Foundation Models to Track, Monitor and Predict SARS-CoV-2 Variants of Concern,*** Dr. Arvind Ramanathan, Argonne National Labs

4:30 – 4:50PM ***Development of a Salmonella enterica Virulence Database and Associated Bioinformatics Analysis,*** Dr. Jing Han, National Center for Toxicological Research, US FDA

4:50 – 5:10PM ***Real-time Genomic Surveillance for Infection Prevention and Antimicrobial Stewardship,*** Dr. Se-Ran Jun, University of Arkansas for Medical Sciences

5:10 – 5:30PM ***Wastewater Surveillance and Pathogen Detection***, Dr. Camila Silva, National Center for Toxicological Research, US FDA

5:30 – 6:00PM ***Break***

6:00PM – 8:00PM ***Poster Session***

DAY 2 – MARCH 14, 2023

8:30AM – 10:30AM ***Session 3 and Session 4 (Parallel)***

SESSION 3: AI IN HEALTHCARE

Co-Chairs: Drs. David Ussery & Jonathan Bona, University of Arkansas for Medical Sciences

Session Overview:

This session includes six lectures from Arkansas scientists that are involved in using machine-learning methods for understanding medical and healthcare data. Three of them are professors from UNIVERSITY OF ARKANSAS FOR MEDICAL SCIENCES, using AI to help in research in cardiovascular medicine, cancer imaging, and drug design. The other three are Ph.D. graduate students from three different graduate programs, using machine learning methods for mining social media for adverse drug reactions, public health, and for ‘crowdsourcing AI solutions’ to challenges in healthcare.

8:30 – 8:50AM ***Using AI in Cardiovascular Healthcare,*** Dr. Subhi J. Al’Aref, University of Arkansas for Medical Sciences

8:50 – 9:10AM ***Mining Reddit for Adverse Drug Reactions***, Catherine Shoults, University of Arkansas for Medical Sciences

9:10 – 9:30AM ***AI in Public Health,*** Brian Delavan, Arkansas Department of Health

9:30 – 9:50AM ***Explainability Analysis of Deep Learning Algorithms in Medical Imaging***, Dr. Fred Prior, University of Arkansas for Medical Sciences

9:50 – 10:10AM ***Crowdsourcing AI Solutions to Challenges in Healthcare***, Jennifer Fowler, Arkansas Economic Development Commission

10:10 – 10:30AM ***AI in Developing Safer Drugs from a Practical Perspective***, Dr. Grover Miller, University of Arkansas for Medical Sciences

SESSION 4: INTEGRATED GENOMICS FOR PRECISION ONCOLOGY

Chair: Dr. Joshua Xu, National Center for Toxicological Research, US FDA

Session Overview:
Tumor initiating and development are driven by molecular events. Over the decades, multiple types of high-throughput omics profiling technologies have been developed and adopted in clinical research. Integrated genomics analysis can help us better understand the mechanisms and phenotypic traits of cancer, tailor treatments and improve clinical outcomes. This session includes four presentations to advance integrated genomics for precision oncology, covering single cell sequencing, best practice in bioinformatics, and real-world data from clinical studies.

8:30 – 8:55AM ***Quality Control and Standardization of Multiomics for Precision Medicine***, Dr. Leming Shi, Fudan University, Shanghai, China

8:55 – 9:20AM ***Spatial Gene Neighborhood Networks by Image-Based Single Cell Genomics***, Dr. Ahmet F. Coskun, Department of Biomedical Engineering, Georgia Institute of Technology & Emory University

9:20 – 9:40AM ***Integrating Multiple Genomic Sequencing Data to Enhance Variant Detection in Tumor Samples***, Dr. Dan Li, Division of Bioinformatics and Biostatistics, National Center for Toxicological Research, US FDA

9:40 – 10:05AM ***SEQC2 – Next Phase with Real World Data***, Dr. Donald J. Johann, Jr., Department of Biomedical Informatics, University of Arkansas for Medical Sciences

10:05 – 10:30AM ***Toward an Improved Risk Stratification for Newly Diagnosed Multiple Myeloma***, Dr. Fenghuang (Frank) Zhan, Department of Internal Medicine, University of Arkansas for Medical Sciences

10:30AM – 11:00AM Break

11:00AM – 12:00PM **Keynote Lecture**

- **Keynote Lecture: Drug Discovery – If You Want to See Different Results Do Things Differently.** Dr. Ruth Roberts, University of Birmingham, UK

12:00PM – 1:00PM Lunch

1:00PM – 2:30PM Session 5 and Session 6 (Parallel)

SESSION 5: DRIVERS OF PUBLIC HEALTH IN THE US POPULATION: WHY VALID SELF-REPORT MEASURES MATTER

Chair: Dr. Samantha Robinson, University of Arkansas, Fayetteville

Session Overview:

This session will serve as a forum for psychometric and public health researchers from two institutions in our state both to share insights from a recent nationally representative survey and to realize the future potential of utilizing validated self-report measures to inform public health initiatives, including those related to mental health. Research related to the conference theme or otherwise relevant will be presented in both 10-minute standard and 15-minute presentation formats, depending upon time allocation for the session. By providing an opportunity for the dissemination of this multi-institutional, multi-disciplinary work of this research group, this session will truly showcase the collaborative power of our Arkansas researchers.

1:00 – 1:30PM ***A Validation Study of Fatalism Scales,*** Dr. Todd Shields, Arkansas State University; Samantha Robinson, University of Arkansas, Fayetteville

1:30 – 1:50PM ***How Adverse Childhood Experiences Influence Fatalism and Mental Health***, Ethan Dennis, University of Arkansas, Fayetteville

1:50 – 2:10PM ***Differential Item Functioning Driven by Intersectionality and Health Risk Behaviors***, Dr. Mary Margaret Hui Cunningham, University of Arkansas, Fayetteville

2:10 – 2:30PM ***Geographical Differences in Self-Report Measures***, Torre (Jake) Darby, University of Arkansas, Fayetteville

SESSION 6: MACHINE LEARNING AND DEEP LEARNING FOR BIG DATA ANALYSIS AND DRUG DEVELOPMENT

Co-Chairs: Drs. Huixiao Hong & Shraddha Thakkar, National Center for Toxicological Research and Center for Drug Evaluation and Research, US FDA

Session Overview:

Currently, artificial intelligence, mainly machine learning and deep learning, is in a Cambrian era with new advanced algorithms being developed and applied across almost all fields. While machine learning and deep learning were developed and applied in the context of other fields, they have all been successfully applied to big data analysis and drug development. To give well-deserved attention to the many facets of applications of machine learning and deep learning to big data analysis and drug development, this session will provide the audience a great lineup of presentations that report most recently progresses in applications of machine learning and deep learning to big data analysis and drug development from well-established scientists in this field from various institutions.

1:00 – 1:30PM ***Deep Learning Based Analysis of Histopathological Images of Breast Cancer,*** Dr. Joe Zhang, University of South Mississippi

1:30 – 1:50PM ***Machine Learning for Repurposing FDA-approved Drugs for COVID-19 Treatment***, Dr. Jie Liu, National Center for Toxicological Research, US FDA

1:50 – 2:10PM ***A Novel Proteogenomics Workflow for Proteoform Detection***, Kanishka Manna, University of Arkansas at Little Rock & University of Arkansas for Medical Sciences

2:10 – 2:30PM ***Genomic Approaches Reveal Mutations in Blood Cells and their Clinical Implications***, Dr. Nuria Mencia-Trinchant, Cornell University

2:30PM – 3:00PM ***Poster Awards and Concluding Remarks***

Speaker Biographies and Abstracts (In alphabetical order by speaker’s last name)

Subhi J. Al'Aref, M.D.
Director, Cardiac CT
Interventional Cardiologist – Assistant Professor of Medicine
University of Arkansas for Medical Sciences (UAMS)
Arkansas, USA

Dr. Subhi Al'Aref is the Director of Cardiac CT at the University of Arkansas for Medical Sciences (UAMS). He is an Interventional Cardiologist and an Assistant Professor of Medicine at the Division of Cardiology, Department of Medicine at UAMS. Dr. Al'Aref's focus includes clinical, educational and research goals. His interests include combining interventional and imaging cardiology, both for clinical and research purposes. He is dedicated to translational research, which takes scientific findings from the lab and translates them to have real-world patient impacts. Dr. Al'Aref also researches artificial intelligence and how it could be used within clinical practice. Dr. Al'Aref has published numerous papers and chapters in textbooks. He has also edited two textbooks, the first on 3D Printing for Cardiovascular Applications and the second on Machine Learning in Cardiovascular Medicine. He is also the principal investigator on an NIH-funded multi-center study on the application of machine learning for cardiac resynchronization therapy (CRT), and is a co-investigatory on an R21 NIH-funded study titled: "Film-like Acoustic Microresonators for Wireless Monitoring of Intracardiac Pressure using Ultrasound". Dr. Al'Aref is also funded by the Bronson Foundation on an investigation that seeks to use artificial intelligence and deep learning for prediction of incident atrial fibrillation on electrocardiography.

Using AI in Cardiovascular Healthcare

Cardiovascular diseases continue to incur significant morbidity and mortality. The traditional clinical workflow involves obtaining a history, performing a physical examination and subsequently utilizing available data (which includes biomarker and imaging results) in order to establish a diagnosis and initiate a therapeutic plan. This traditional workflow suffers from significant limitations, which include inefficient workflow when it comes to data extraction, image result analysis, and integration with patient-level characteristics. In addition, current prognostic modeling has limited utility as a result of poor power to discriminate between short- and long-term adverse events, as well as the usage of limited amount of clinical and imaging data. The integration of artificial intelligence promises to overcome all these limitations. The purpose of the present talk is to give an overview of the current applications of artificial intelligence within cardiovascular healthcare: from automation of coronary artery calcium scoring on computed tomography, to the deployment of AI and deep learning from the prediction of future atrial fibrillation using baseline electrocardiography. The talk will also cover the application of artificial intelligence of the prediction of the response to cardiac resynchronization therapy (CRT) in patients with heart failure.

Namandjé N. Bumpus, Ph.D.
Chief Scientist
Office of the Chief Scientist
U.S. Food and Drug Administration (FDA)
Silver Spring, MD, USA

Dr. Namandjé N. Bumpus was named as the FDA's Chief Scientist on June 30, 2022. The Office of the Chief Scientist supports the research foundation, science, and innovation that underpins the FDA's regulatory mission. It does this through a broad framework that encompasses scientific collaborations, laboratory safety, the transfer of FDA inventions to the private sector, scientific integrity in FDA policy- and decision-making, the professional development of regulatory scientists, and its core research component—the FDA's National Center for Toxicological Research—which generates the vital data that the FDA requires for its regulatory decision-making and development of sound regulatory policy.

Before joining the FDA, Dr. Bumpus was the E.K. Marshall and Thomas H. Maren Professor and chair of the Department of Pharmacology and Molecular Sciences at the Johns Hopkins University School of Medicine. She served previously as associate dean for basic research in the Johns Hopkins University School of Medicine. Dr. Bumpus' research has focused on drug metabolism, pharmacogenetics, bioanalytical chemistry, and infectious disease pharmacology. Dr. Bumpus joined the faculty at Johns Hopkins in 2010 as an assistant professor. She earned a bachelor's degree in biology at Occidental College in 2003, a doctorate in pharmacology at the University of Michigan in 2007 and completed a postdoctoral fellowship in molecular and experimental medicine at The Scripps Research Institute in La Jolla, CA in 2010.

Dr. Bumpus currently serves as president-elect of the American Society for Pharmacology and Experimental Therapeutics. She previously served as chair of the NIH Xenobiotic and Nutrient Disposition and Action study section.

Her many honors include the Leon I. Goldberg Award from the American Society for Clinical Pharmacology and Therapeutics, the James Gillette Award from the International Society for the Study of Xenobiotics, the John J. Abel Award in Pharmacology from the American Society for Pharmacology and Experimental Therapeutics and the Presidential Early Career Award for Scientists and Engineers, which is the highest honor bestowed by the United States government on early career scientists and engineers. Dr. Bumpus is an elected fellow of the American Association for the Advancement of Science. She became a Member of the National Academy of Medicine, Class of 2022, one of the highest honors in the fields of health, science and medicine.

Ahmet F Coskun, Ph.D.
Bernie Marcus Early-Career Professor
Wallace H. Coulter Department of Biomedical Engineering
Georgia Institute of Technology & Emory University
Atlanta, Georgia

Dr. Coskun is currently a Bernie-Marcus Early-Career Professor of Biomedical Engineering at Georgia Institute of Technology and Emory University. Dr. Coskun is a systems biotechnologist and bioengineer, working at the nexus of multiplexed cell imaging and quantitative tissue biology. Dr. Coskun directs an interdisciplinary research team at the Single Cell Biotechnology Laboratory, an interdisciplinary program that is strategically positioned for multiparameter imaging one cell at a time by spatial context and function. Dr. Coskun holds 5 issued patents and is also the co-author of >40 peer-reviewed publications in major scientific journals. Previously, Dr. Coskun was an Instructor at Stanford University. Dr. Coskun received his postdoctoral training from the California Institute of Technology. He holds a Ph.D. degree from the University of California, Los Angeles. His research has been supported by the National Institutes of Health (NIAID and NCI), Wellcome LEAP, Burroughs Wellcome Fund (CASI), NSF CMaT, American Lung Association, American Cancer Society IRG, Multi-cellular engineered living systems (M-CELS), and Regenerative Medicine Center. He is a recipient of the Student Recognition of Excellence in Teaching: Class of 1934 CIOS Award. In addition, he leads outreach programs to engage K12 students and undergraduate students through BioCrowd Studio, an innovative crowd-sourcing program bringing together interactive virtual media, distributed biokits, and collaborative spatial discovery.

Spatial Gene Neighborhood Networks by Image-Based Single Cell Genomics

The spatial organization of cells in tissues and subcellular networks provides a quantitative metric for determining health and disease states. Single-cell analyses of molecular profiles with in-situ detection methods dissect spatial heterogeneity of distinct cell types. Such detailed cellular digital maps shed light on the spatial regulation mechanisms of many disorders. The next challenge in spatial biology is to link the cellular functional responses to the cell identities and phenotypes in their native three-dimensional (3D) environments. To achieve this important goal, image-based multiparameter molecular profiling has the potential to decode high-dimensional dynamics of signaling and metabolism at the subcellular and molecular levels in complex tissues and organs. In this talk, I will introduce multiplex imaging modalities (genomics and beyond) and a novel spatially resolved gene neighborhood network (SpaGNN) concept to decipher the spatial and temporal decision-making of single cells at macromolecular resolution in engineered organoids and human tissues for subcellular and cellular precision oncology. Automated machine learning algorithms in this single-cell big data impact biomedical practice and clinical care.

Mary Margaret Hui Cunningham, Ed.D.
Instructor, Data Analysis
Department of Management, Sam M. Walton College of Business
University of Arkansas
Fayetteville, Arkansas, USA

Dr. Mary Margaret Cunningham is an Instructor of Data Analysis in the Sam M. Walton College of Business at the University of Arkansas. She holds a doctorate in higher education administration and is completing another doctorate in educational statistics and research methods. Her research areas include differential item functioning, QuantCrit, and student success.

Differential Item Functioning Driven by Intersectionality and Health Risk Behaviors

Informed by an intersectional understanding of differences in demographic information (including age, gender, race, education, ability, and having a child under 18), this session will evaluate responses to a fatalism scale.

Detection of Differential Item Functioning (DIF), including DIF attributable to complex combinations of observed covariates, is relevant in such a context. However, DIF detection techniques often compare pre-specified groups, with numeric covariates arbitrarily split (often at the median).

Strobl, Kopf, and Zeileis (2015) proposed Rasch trees as a new method for DIF detection. The methodological framework was then extended to polytomous items by Komboz, Strobl, and Zeileis (2018). The tree-based, latent class approach, utilizing model-based recursive partitioning, serves as both a model test and a global test for DIF. Rasch trees can automatically detect groups exhibiting DIF without needing to pre-specify groups or numeric cutpoints. This approach splits the sample at the strongest point of parameter change. In addition to improved splitting, Rasch trees can detect DIF with more than one covariate, highlighting the intersectionality (i.e., the interaction) of variables rather than detecting DIF with a singular covariate. This session will demonstrate how Rasch tree DIF detection complements intersectional variables.

Jake Darby
Honors College Fellow
Department of Mathematical Sciences
University of Arkansas
Fayetteville, Arkansas, USA

Jake Darby is an Honors College Fellow at the University of Arkansas studying in both the Department of Mathematical Sciences (MASC) and the Computer Science and Computer Engineering Department (CSCE). He is also currently studying to become an actuary and has passed the first two certification exams (Probability and Financial Mathematics). As part of Dr. Robinson's research group, he has worked extensively with large scale survey data and his current research focuses on exploring geographical differences in instruments at the test and item level.

Geographical Differences in Self-Reported Measures

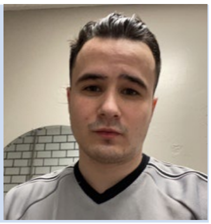
Emergence of patient-focused measures and sustained interest in large-scale surveys (e.g., Demographic and Health Survey [DHS]) has led to broad use of psychometric models (e.g., Item Response Theory [IRT]) in health policy to assess latent (i.e., unobserved) traits such as quality of life, health-related attitudes, and knowledge.

Detection of differences attributable to geographic location, is relevant in such a context. Current work utilizes national survey data to detect and explore geographical differences in item functionality as well as differences in overall trait levels for various quality of life, mental health, and fatalism measures.

Rasch trees, a differential item functioning (DIF) method based upon model-based recursive partitioning, provided evidence that individual items on certain scales differed based upon geographic location. Additionally, Analysis of Variance (ANOVA) methods revealed that geographic location had a significant effect on overall trait levels.

These preliminary findings reveal features that contribute differentially to quality of life, mental health, and fatalism levels and also provide cross-sectional snapshots of current regional differences in the United States.

Ethan Dennis
Research Assistant
Department of Psychological Science
University of Arkansas
Fayetteville, Arkansas, USA



Ethan Dennis is a Research Assistant for the Treating Emotion And Motivational Processes Transdiagnostically (TEMPT) Lab in the Department of Psychological Science at the University of Arkansas. He has also assisted with statistical programming in projects across multiple disciplines at the University of Arkansas. His current research examines a variety of facets of emotion dysregulation and associated psychopathologies by utilizing ecological momentary assessment (EMA). His most recent and notable research projects have included 1) an analysis of the effect of affect lability and self-criticism lability on affect intensity, valence, and levels of self-criticism; 2) the extent to which overcontrolled or undercontrolled tendencies predict momentary fluctuations in suicidal ideation; and 3) the adaptation and validation of a psychometric scale concerning motives behind vaping derived from a scale measuring smoking motives.

How Adverse Childhood Experiences Influence Fatalism and Mental Health

Adverse Childhood Experiences (ACEs) remain one of the most significant public health issues in the United States. Elucidating the mechanisms through which the experience of ACEs may influence mental health outcomes in the future will better prepare clinicians and researchers alike to formulate treatments through which these mechanisms may be taken into account. The current study seeks to:

- Examine the directionality of the relationship between ACEs and fatalism
- To see how these constructs both interact and operate individually to predict mental health outcomes such as depression and anxiety

Jennifer Fowler
Project Director & Principal Investigator, Arkansas NSF EPSCoR
Arkansas Economic Development Commission
PhD Candidate, Molecular Biosciences
Arkansas State University

Jennifer Fowler is the statewide program director of Arkansas NSF EPSCoR, and principal investigator for the current Track-1 project, Data Analytics that are Robust and Trusted (DART, OIA-1946391). Jennifer is an Arkansas native and is passionate about research and creating more opportunities for learners. After attending the Arkansas School for Mathematics, Sciences, & the Arts (ASMSA), she obtained her bachelor's degree in Biology from the University of Arkansas. Prior to her promotion to PI/PD, she served for 7 years as the Director of Education, Outreach, & Diversity for Arkansas NSF EPSCoR, where she implemented a variety of programs and activities to broaden participation in science, technology, engineering, and math (STEM). She is currently pursuing her PhD in Molecular Biosciences at Arkansas State University with an emphasis in machine learning. Her thesis research involves evaluating machine learning approaches to identify possible genetic markers for types of cancer. She also serves as Director of Partner Engagement for National AI Campus, and on the board for Girl Scouts- Diamonds of Arkansas, Oklahoma, and Texas, and the ASMSA Board of Visitors.

Crowdsourcing AI Solutions to Challenges in Healthcare

In this talk, Jennifer will provide a brief overview of Arkansas EPSCoR and introduce the current Track-1 project, Data Analytics that are Robust and Trusted (DART). She will highlight some examples of DART-supported work in AI for healthcare and discuss the current Federal funding landscape for this type of work.

She will then provide an overview of her research- a novel algorithm to compare differential gene expression profiles and harness the large amounts of publicly available data to identify candidate genes for targeted therapy for cancer and present a broader discussion of the future of personalized medicine/precision oncology enabled by AI.

Finally, she will share her experience from the AI for healthcare competitions that she has participated in through AI Campus (Kidney Tumor Segmentation challenge, COVID-19 DREAM Challenge, and Long COVID challenge) including a broader discussion of crowdsourcing solutions to challenging questions in healthcare and related data governance/privacy issues.

Jing Han, Ph.D.
Division of Microbiology
National Center for Toxicological Research (NCTR)
U.S. Food and Drug Administration (FDA)
Arkansas, USA

Dr. Jing Han is a research microbiologist in the Division of Microbiology at FDA's National Center for Toxicological Research (NCTR) in Jefferson, Arkansas. Her main research interests are in the fields of antimicrobial resistance, pathogenesis of foodborne pathogens, and genetic characterization of enteric bacteria using molecular techniques. Her research projects at NCTR include: 1) Characterization and Assessment of Potential Targets as Antivirulence Drug for Avian Pathogenic *Escherichia coli* (APEC); 2) Development of improved databases for *Salmonella* virulence, gene identification, and plasmid characterization and development of the analytical tools for data analyses; 3) Genetic characterization of antimicrobial resistance and associated genetic factors in *Salmonella* serovars associated with food animals and invasive human infections; 4) Sequencing and functional analysis of plasmids isolated from multi-antimicrobial resistant bacteria; 5) Evaluation of the relative selective potential of antimicrobial agents to trigger the dissemination of antimicrobial resistance and virulence factors to susceptible *Salmonella*; 6) Investigation of microbial populations in different tobacco products. Dr. Han has published more than 50 papers and book chapters.

Development of a *Salmonella enterica* Virulence Database and Associated Bioinformatics Analysis

Background: *Salmonella enterica* is a major foodborne pathogen. Within the species there appears to be significant genetic diversity that leads to differences in *Salmonella*'s ability to infect humans and animals, and impact disease severity. Whole-genome sequencing (WGS) provides rich data on the genetics of virulence and antimicrobial resistance; however, improved bioinformatics tools are required to fully utilize the wealth of WGS data. There is a need for improved tools to rapidly detect and characterize virulence in strains that may be associated with outbreaks. This study was undertaken to develop a comprehensive virulence-factor database and computational tools to determine the genes present within strains that may impact pathogenicity.

Methods: A comprehensive list of virulence genes was generated based on existing databases and literature searches. DNA sequences and corresponding metadata were imported into the database. Matching algorithms were developed to predict virulence genes present in WGS data, along with analysis tools for virulence gene comparisons. About 50,000 *Salmonella* WGS from 14 different serotypes were downloaded from NCBI and analyzed using our database and BioNumerics.

Results: A virulence gene database containing ~500 virulence or putative virulence genes was constructed. Simple matching algorithms and virulence gene comparison tools were developed and deployed into the database. Phylogenetic and comparative analyses of the virulence gene profiles of isolates showed that strains of same serotypes tended to cluster, yet differences among isolates in serotypes could be assessed.

Conclusion: Our virulence database can identify potential virulence factors present in *Salmonella* and allow comparison of virulence profiles among different isolates.

Donald J. Johann, Jr., MD, MSc, FACP
Professor, UAMS College of Medicine
Director, UAMS Genomics Facility
University of Arkansas for Medical Sciences (UAMS)
Arkansas, USA

Dr. Johann is a physician/scientist (medical oncologist), Professor at UAMS and the inaugural Director of the UAMS Genomics Sequencing Facility. His scientific focus concerns developing more effective therapies and molecular diagnostics for patients with cancer. The characterization of genetic-based alterations in model systems and bio-liquids using advanced molecular profiling and high-throughput technologies are active and significant interests, and especially for lung cancer. Areas of emphasis include: drug development utilizing model systems, cancer biology, bioinformatics, advanced molecular profiling (NGS), & advanced tissue microdissection (eg, LCM). As the Director of the UAMS Genomics Sequencing Facility, now provided are state-of-the-art capabilities for solid tissues (including FFPE) and bio-liquids, eg, WGS (including ultra-low-pass), panels using Universal Molecular Identifiers (UMIs), 10x single cell RNA-seq, advanced epigenetics & bioinformatics.

Prior to joining UAMS he was an investigator at the NIH/NCI Center for Cancer Research in Bethesda, MD. Prior to medical school, he worked as an engineer for the Sperry/Unisys Corp for six years. Rising thru the ranks, he directed a group of five engineers on R&D projects involving systems level software (OS, compilers), advanced avionic software and systems integration with innovative instrumentation.

SEQC2 – Next Phase with Real World Data

Background

The Sequencing Quality Control Phase 2 (SEQC2) consortium is an international group, composed of members from academia, industry, and led by the FDA. The aims of this group are the development of best practices, protocols, and quality metrics for NGS-based diagnostic assays, which support precision medicine and regulatory science. It is intended to support cancer research, such as the development of more sensitive diagnostic tests along with research that has the potential to transform the scientific field and improve patient outcomes. The purpose of this study is to introduce a new phase of SEQC2 focused on *Real World Data* (RWD).

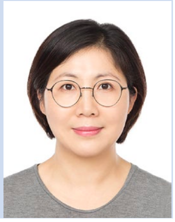
Results

A targeted panel consisting of over 500 genes relevant to human cancers has been designed for DNA and RNA (two separate panels, same gene targets) and will utilize Unique Molecular Identifiers (UMIs) to improve the overall fidelity of the assays by reducing the likelihood of false positive findings. Targeted assays will be run on matched samples from real cancer patients consisting of tissue from solid tumors (DNA and RNA) and circulating tumor DNA (ctDNA). The ctDNA samples have been obtained from lung cancer patients in the following time frames: pre-operative, post operative, and longitudinally at standard-of-care visits. Agilent Universal RNA Reference sample (UHRR) and SEQC2 reference materials developed by the Oncopanel Sequencing Working Group will be used as a special "clinical sample" and included as "batch controls" during the NGS runs.

Conclusions

The SEQC2 consortium sponsored the liquid biopsy proficiency study that was successfully conducted by the Oncopanel Sequencing Working Group and published in Nature Biotechnology. Advanced reference materials were developed by the working group and used in this study as ground truth. Applying insights gained from the proficiency study to RWD is the next logical step and will be an important collaborative endeavor.

Se-Ran Jun, Ph.D.
Assistant Professor
Department of Biomedical Informatics
University of Arkansas for Medical Sciences
Little Rock, Arkansas USA

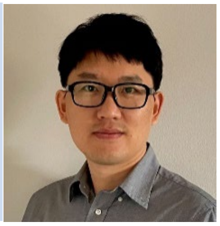


Dr. Se-Ran Jun is currently an assistant professor of Translational Bioinformatics at the Department of Biomedical Informatics, University of Arkansas for Medical Sciences (UAMS). She obtained her Ph.D. in Applied Mathematics from Pohang University of Science and Technology (POSTECH), was a Senior Engineer in Computational Mathematics group within the Computational Science and Engineering Division at Samsung Advanced Institute of Technology (SAIT) and was a Research Staff at UT-ORNL Joint Institute for Computational Sciences, Knoxville and University of California, Berkeley. Her research interests include Translational Bioinformatics, Computational Genomics, Computational Microbiome, Systems Biology with emphasis in Cancer and Infectious Disease. Her research areas are genomic and microbiome epidemiology at the intersection of next generation sequencing, big data science, biocomputing, and biomedical sciences. Her lab has a long-standing interest in microbial genomics and microbiome research. Ongoing studies are focused on (1) genomics of drug resistance, (2) bacterial infection and transmission within communities or the hospital environment, and (3) microbiome and resistome dynamics. Currently, her lab is working on genomic surveillance for infection prevention and antimicrobial stewardship using real-time sequencing technology and dynamics of antibiotic resistome between farms and supply chains. Dr. Jun has published about 50 papers and book chapters, and is involving 4 grants as a PI and 5 grants as a co-I. More information about her group and research interests can be found at https://uams-triprofiles.uams.edu/profiles/display/3642362.

Real-time Genomic Surveillance for Infection Prevention and Antimicrobial Stewardship

Whole genome sequencing provides detailed information about pathogens, and further data-rich understanding of bacterial antibiotic resistance. Integration of AR genomic data into routine clinical practice offers new and promising opportunities to significantly improve patient outcomes, thus making hospitals safer. Success of this depends on our goals of development of a system with 1) fast enough turnaround time and 2) the discriminatory power to impact antibiotic stewardship decision and to assist in breaking transmission routes to improve interventions made by infection prevention teams. Thus, we have demonstrated early success of our real-time genomic surveillance through a proof-of-concept model using isolates of vancomycin resistant *Enterococcus faecium*, a pathogen of high clinical consequence collected from cancer patients at UAMS. Our study confirmed the prospect of genetic AR data enhancing antibiotic stewardship and the power of integration of genomic data into the Electronic Medical Record for “precision infection control.” With this success, our team is expanding our study to the ESKAPE group pathogens.

Dan Li, Ph.D.
Division of Bioinformatics and Biostatistics
National Center for Toxicological Research (NCTR)
U.S. Food and Drug Administration (FDA)
Arkansas, USA

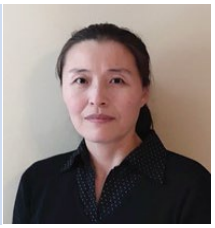


Dr. Dan Li is a Visiting Scientist in the Division of Bioinformatics and Biostatistics at the FDA's National Center for Toxicological Research (NCTR/FDA). He has a diverse background in computer science and bioinformatics. His current research interests include genomic sequencing, single-cell NGS, long non-coding RNA, data analysis, and data mining. Dr. Li is involved in several projects that develop bioinformatic methodologies and standards to support FDA research and regulation, such as the Microarray and Sequencing Quality Control phase II (MAQC2/SEQC2). His work mainly focuses on targeted genomic sequencing data, oncology panels, tumor mutational burden, and the detection of genomic variants and fusions. In addition, Dr. Li participates in the development of new tools and methods to improve the storage, sharing, collection, and analysis of pharmacology and toxicology review data, and to support FDA review and research on pharmacogenomics.

Integrating Multiple Genomic Sequencing Data to Enhance Variant Detection in Tumor Samples

Genetic mutations that impact protein function are a major cause of human disease, but the impact and penetrance of these mutations may depend on alterations to protein binding sites or transcription within a gene. RNA sequencing is increasingly being used to interpret the impact of mutations on gene expression and splicing, in addition to screening for mutations using cancer DNA panels. However, there has not been a thorough performance assessment of RNA panels, nor have data analytics been used to screen for transcribed variants in addition to mutations identified in cancer DNA panels. We performed matched targeted DNA and RNA sequencing on a set of reference samples with pre-identified true variants in order to evaluate the detection and interpretation of genetic mutations transcribed within genes. We also included non-captured whole transcriptome RNA-seq data in our comprehensive comparison. Our focus was on the metrics of reproducibility, recall, and precision for variant detection in tumor, normal, and mixed samples. We also investigated the variant allele frequency (VAF) of the same variants in titrated samples to measure the performance of different sequencing methods. As a result, we observed good performance for both DNA-seq and RNA-seq data. By integrating with (targeted) RNA-seq data, we can confirm and highlight expressed variants in order to prioritize during genetic diagnosis and analysis. We recommend using RNA-seq as a complement to DNA-seq to enhance variant detection in tumor samples.

Jie Liu, Ph.D.
Staff Fellow, Division of Bioinformatics and Biostatistics
National Center for Toxicological Research (NCTR)
U.S. Food and Drug Administration (FDA)
Arkansas, USA

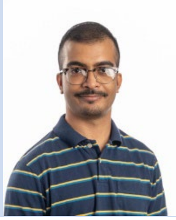


Dr. Jie Liu is a staff fellow of Division of Bioinformatics and Biostatistics at FDA's National Center for Toxicological Research (NCTR/FDA). Dr. Liu's specialized research focuses on the development of machine learning models and databases for safety evaluation and risk assessment. Currently, Dr. Liu works on the development of machine learning models for in vivo toxicity prediction, repurposing FDA-approved drugs for COVID-19 treatment and predicting opioid receptor binding activity for assisting the development of opioid drugs. Dr. Liu has published over 20 peer-reviewed papers and book chapters.

Machine Learning for Repurposing FDA-approved Drugs for COVID-19 Treatment

COVID-19 is a global pandemic with millions of people infected. Although US FDA approved several drugs for the treatment of COVID-19, effective COVID-19 treatment drugs are in urgent need. The main protease of SARS-CoV-2 is a major target for COVID-19 drugs. In this study, we applied machine learning for repurposing FDA approved drugs that could bind the main protease of SARS-CoV-2 as potential candidates for the treatment of COVID-19. We collected 3D structures of the main protease of SARS-CoV-2 bound with ligands from the protein data bank. The 372 ligands were separated from the structures and were used as binders for training. We then curated some 400 compounds experimentally tested in SARS-CoV-2 main protease binding assays from the literature. Of the curated non-binders, 188 were used for training. The rest compounds (both binders and non-binders) were used for testing. We curated 1284 FDA-approved drugs from diverse sources including drug labeling documents for identification of SARS-CoV-2 main protease binders for repurposing. Random forest algorithm was used for constructing predictive models based on molecular descriptors calculated using Mold2 software. Model performance was evaluated using 500 iterations of 5-fold cross validations and the testing data set. The random forest models showed 84.2% and 78.6% prediction accuracy in the 5-fold cross validations and the testing, respectively. The random forest model constructed from the whole training data set was used to predict SARS-CoV-2 main protease binders as potential candidates for repurposing to COVID-19 treatment. Our results demonstrate that machine learning could be an efficient method for drug repurposing, and thus accelerate the drug development targeting SARS-CoV-2.

Kanishka Manna, M.Sc.
Ph.D. Candidate, Joint Bioinformatics Program
University of Arkansas at Little Rock (UALR) &
University of Arkansas for Medical Sciences (UAMS)
Arkansas, USA



Kanishka Manna is a Ph.D. Candidate in the Joint Bioinformatics Program, at the University of Arkansas at Little Rock (UALR) and the University of Arkansas for Medical Sciences (UAMS). Currently, his Ph.D. dissertation project involves the efficient identification of protein isoforms originating from different events such as alternative splicing by developing a proteogenomics pipeline for novel peptide detection among multi-omics samples. Additionally, the project also involves accurate annotation for identifying the newly identified peptide sequences. Presently, he is being mentored by Dr. Stephanie Byrum, Ph.D., Associate Professor and director of bioinformatics shared resources at the University of Arkansas for Medical Sciences (UAMS). Mr. Manna's overall research goal is to discover therapeutic targets for different diseases using a systems biology approach.

A Novel Proteogenomics Workflow for Proteoform Detection

Multi-omics experimental approaches such as proteogenomics are becoming a mainstream practice in cancer biology underlining the need to design new integrative techniques and applications to enable the multi-scale characterization of different types of cancer for targeted therapy. Proteogenomics incorporates techniques to integrate genomic, transcriptomic and proteomic data for efficient identification of novel proteins. The explosion of biological big data has outperformed our capability to manipulate, mold and leverage for novel insights with efficacy. Additionally, there are a multitude of challenges when applying proteogenomics analyses, such as improper integration of each omics data type, missing structural annotations, inefficient identification of splice-junctions, novel genes, etc.

This research has two primary aims. In the first aim, the plan is to build a reproducible, adaptive and automated proteogenomics pipeline and custom protein sequence database where changes in the DNA sequence of genes (genetic driver mutations), obtained from the genomic and transcriptomic data will be integrated concurrently with the proteomics data for efficient and precise identification of novel protein isoforms.

The second aim involves building a reproducible bioinformatics workflow that will accurately annotate and curate a consensus protein reference sequence database, along with its associated coding DNA sequence. This reference database will include information pertaining to their structural annotations from any available reference genomes.

Finally, through these aims, the newly developed proteogenomics pipeline will be benchmarked under a variety of parameter settings. The proteogenomics pipeline will be applied to create and curate a custom proteogenomics database in order study the patient derived xenograft (PDX) model system. Specifically focusing on the identification of unique resistance markers, such as MAPKi, to investigate alterations and putative therapeutic vulnerabilities of these markers.

Grover P. Miller, Ph.D.
Professor
Departments of Biochemistry and Molecular Biology and of Biomedical Informatics
University of Arkansas for Medical Sciences
Arkansas, USA

Dr. Grover P Miller is a Professor in the Department of Biochemistry and Molecular Biology with a secondary appointment in Biomedical Informatics at the University of Arkansas for Medical Sciences. His research focuses on assessing the mechanistic roles for metabolism, bioactivation, and clearance of molecules in pharmacology and toxicology. In practice, his group leverages powerful computational, analytical, and biochemical tools to predict, identify, and quantitate small molecules including drugs, dietary molecules, and pollutants during metabolism and correlate findings to biological activity and *in vivo* outcomes such as drug-induced liver injury. Individual projects aim to (1) determine metabolic mechanisms, efficiencies, and fluxes for activation, deactivation, and elimination of molecules, (2) identify metabolite biomarkers in humans and animal models for correlating *in vitro* findings to *in vivo* outcomes and leveraging their diagnostic, theragnostic, and prognostic potential, and (3) develop computational models for drug metabolism and bioactivation contributing to adverse drug events to make drugs safer for clinical use. After over 30 years in the field, his efforts have been recognized through 63 peer-reviewed manuscripts and 13 reviews along with roles as Chair (4) and Reviewer (38) on research study sections, membership on journal editorial boards (5), and session chair (10) and organizer (9) for research meetings. Over time, his research expanded from detailed *in vitro* metabolic studies to metabolite profiling for translational studies and development of experimental and computational models of metabolism, structure, and reactivity that were made possible through strong, interdisciplinary collaborations.

AI in Developing Safer Drugs from a Practical Perspective

AI models of metabolism by cytochromes P450 or general reaction types are revolutionizing drug discovery and development; however, their effective incorporation in the workflow poses challenges due to contrasts in design relative to current, traditional benchtop experimental methods. We have learned valuable, practical lessons on read outs, predictability, and scaling of deep neural metabolism models while coupling their use to experimental methods for assessing model performance in applications. First, our models generate a read-out that does not translate to commonly used metabolic parameters. Nevertheless, we estimated terbinafine metabolic flux through competing N-dealkylation pathways using model reaction probabilities. This approach successfully distinguished between major and minor terbinafine metabolic pathways. Second, model accuracy depends on whether training information is applicable to molecules of interest. As a practical test, we explored the effectiveness of our epoxide model to predict the impact of substituents on thiazole bioactivation using sudoxicam as a template. The model performed well with alkyl substituents and others but poorly with halogens indicating its strengths and weaknesses. Modelling can also predict unexpected bioactivations as we reported for conversion of isoxazole-containing bromodomain inhibitors into reactive extended quinones. Third, model read outs scale relative to a specific training population but not to other models, which precludes their collective use in decision-making. As an alternative, we used a generalized oxidative bioactivation model to predict families of drugs prone to reactive quinone formation that could explain reported drug-induced liver injuries. We followed up experimentally characterizing bioactivation pathways for diphenylamine nonsteroidal anti-inflammatory drugs. In the process, we demonstrated how modelling effectively guided experimental studies despite limits in the granularity of predictions. Taken together, insights gained from these practical applications aided in understanding how models can provide insights on drug metabolism and bioactivations as well as identifying ways to improve the models to realize their potential.

Amanda Novack, MD
Medical Vice President of Quality and Safety
Baptist Health
Arkansas, USA

Dr. Amanda Novack is an Arkansas native who graduated from Hendrix College and earned her medical degree from University of Arkansas for Medical Sciences. She completed a residency in Internal Medicine and a fellowship in Infectious Diseases at UAMS as well. She served as faculty there for two years, with an emphasis on expediting discharges and optimizing home IV antibiotics. She worked with the Arkansas Department of Health providing statewide education about antimicrobial stewardship and infection prevention, and eventually served on the medical staff of 15 hospitals, providing consultation for hospital policy, as well as individual patients through telemedicine and in-person visits. In November 2019, she was named Medical Director of Infection Prevention for Baptist Health, and led much of the system's response to the COVID-19 pandemic. She was recently named Medical Vice President of Quality and Safety at Baptist, and she continues to live in Little Rock with her family and too many dogs.

Reduction of Central-Line Associated Bloodstream Infections, with Less Staff and More Patients

A **Central-Line Associated Bloodstream Infection (CLABSI)** is a hospital acquired infections that increase cost, length of stay, and morbidity in hospitalized patients. Prevention of these infections requires a multipronged approach, which includes daily maintenance of the line itself, along with minimizing use of these high-risk IV catheters whenever possible. During the COVID-19 pandemic, the number of central lines among inpatients averaged 100/day, and our resources did not allow for intense scrutiny of each individual line. Using strategic data analysis through our electronic medical record (EMR), we were able to focus efforts on the highest-risk central lines, and decrease CLABSIs despite higher census and fewer resources.

- **Root cause analysis** of the last 50 CLABSIs revealed that the most common failures fell in these categories: central lines that no longer met criteria for placement, patients who had other sources of infections not identified in a timely manner, and patients who were cultured as they were actively dying.
- **Reports were built in our EMR** to identify patients who fell into these categories, and allowed for strategic, efficient interventions
- **High risk patients** received extra rounding to review line maintenance, and extra investigation around other potential sources of infection
- **Between December 2019 and October 2022,** we were able to reduce the standardized infection ratio (actual CLABSIs divided by CLABSIs predicted by the National Healthcare Safety Network) from 1.723 to 0.410
- **CLABSIs among ICU patients** decreased from a total of 27 in 2020, to a total of 5 in 2022.

Aaron M. Novotny, Ph.D.
Director, Healthcare Economics and Outcomes
Arkansas Blue Cross Blue Shield (ABCBS)
Arkansas, USA

Dr. Aaron M. Novotny is Director of Healthcare Economics and Outcomes at Arkansas Blue Cross Blue Shield. He serves as an adjunct professor at the University of Arkansas Sam Walton College of Business and at the University of Arkansas for Medical Sciences Faye Boozeman College of Public Health. Dr. Novotny has published papers in Social Desirability and in Developmental Economics. His division at ABCBS concentrates on methodologies and research through advanced statistical techniques to support healthcare intervention strategies and optimize outcomes. The most visible projects from his group are (1) Birds of a Feather: analysis on how health shocks for a particular family member shift healthcare utilization for other household members; (2) understanding the social determinants of health on utilization patterns and disease progression rates; (3) examining the impact behavioral health services have on physical health cost of care and chronic condition maintenance; and (4) maternal analysis to understand birth outcomes for underserved populations to reduce infant and maternal mortality in AR.

Elizabeth Parker, PhD, MPH
Sr. Director, Principal Data Scientist
Arkansas Blue Cross Blue Shield (ABCBS)
Arkansas, USA

Dr. Elizabeth Parker has worked in the healthcare industry conducting research and clinical trials in various arenas of healthcare. After a six-year stint in the private retail sector, she has made her way back to healthcare as the Principal Data Scientist for Arkansas Blue Cross & Blue Shield. She also serves as a professor of biostatistics, epidemiology, clinical research, and leadership at George Washington University Medical School. Additionally, she serves on the Advisory Board for The Forge Institute and Chair of the Board for The Museum of Discovery.

Birds of a Feather: Health Events lead to Changes in Household Healthcare Utilization

During the height of the pandemic, we observed full ERs limiting access to individuals, quick transmission of the Covid-19 virus among families and community members, and wholesale shifts in how we utilize mental health services from in-person to virtual care. These changes in the healthcare landscape demonstrated that members' health impact one another. We explore this notion by examining household healthcare utilization after a large-scale shift by a single member. By understanding the impact on the surrounding family members (many which have limited to no data), we can begin to understand, forecast, and predict future utilization and identify community health patterns to improve health across Arkansas. Moreover, data feature construction can leverage these factors to improve member health outreach and outcomes models.

Fred Prior, Ph.D.
Distinguished Professor and Chair
Department of Biomedical Informatics
University of Arkansas for Medical Sciences
Arkansas. USA

Dr. Fred Prior is Distinguished Professor and Chair of the Department of Biomedical Informatics and Professor of Radiology at the University of Arkansas for Medical Sciences (UAMS). Dr. Prior's research interests include cancer informatics, radiomics, and neuroimaging informatics. He serves as principal investigator and director of the US National Cancer Institute's Cancer Imaging Archive project and is the lead PI of an NCI ITCR team exploring the integration of radiomics and pathomics. In 2021 Dr. Prior's team joined a consortium of European colleagues to successfully compete for a Horizon 2020 award from the EU to develop a platform for distributed data management and machine learning to advance precision medicine in oncology. Dr. Prior directs informatics efforts for the UAMS Translational Research Institute and Pediatric Clinical Trial Network. He is an associate editor of several leading scientific journals, and a reviewer for numerous scientific and engineering journals as well as U.S. and European funding agencies. He is the author of over 150 scientific publications and holds 6 US and international patents.

Explainability Analysis of Deep Learning Algorithms in Medical Imaging

Machine learning, especially deep learning is being applied in many areas of current biomedical research. There is growing concern that results produced by these AI techniques are inconsistent, not necessarily biologically relevant and that "black box" algorithms are not trustworthy. Explainability analysis employs techniques that can deconstruct the machine decision-making process and reproduce the learning and knowledge extraction processes used the AI algorithm. Such analyses serve to verify system functionality, help to establish trust, and maybe identify previously unknown characteristics that can lead to new understanding of disease mechanisms. This talk will introduce key concepts of explainability analysis in cancer imaging research.

Arvind Ramanathan, Ph.D.
Computational Science Leader
Data Science and Learning Division
Argonne National Laboratory
Illinois. USA

Dr. Arvind Ramanathan is a computational biologist in the Data Science and Learning Division at Argonne National Laboratory and a senior scientist at the University of Chicago Consortium for Advanced Science and Engineering (CASE). His research interests are at the intersection of data science, high performance computing and biological/biomedical sciences. His research focuses on three areas focusing on scalable statistical inference techniques: (1) for analysis and development of adaptive multi-scale molecular simulations for studying complex biological phenomena (such as how intrinsically disordered proteins self-assemble, or how small molecules modulate disordered protein ensembles), (2) to integrate complex data for public health dynamics, and (3) for guiding design of CRISPR-Cas9 probes to modify microbial function(s). He has published over 50 papers, and his work has been highlighted in the popular media, including NPR and NBC News. He obtained his Ph.D. in computational biology from Carnegie Mellon University and was the team lead for integrative systems biology team within the Computational Science, Engineering and Division at Oak Ridge National Laboratory. More information about his group and research interests can be found at http://ramanathanlab.org.

Using Foundation Models to Track, Monitor and Predict SARS-CoV-2 Variants of Concern

We seek to transform how new and emergent variants of pandemic-causing viruses, specifically SARS-CoV-2, are identified and classified. By adapting large language models (LLMs) for genomic data, we build genome-scale language models (GenSLMs) which can learn the evolutionary landscape of SARS-CoV-2 genomes. By pre-training on over 110 million prokaryotic gene sequences and fine-tuning a SARS-CoV-2-specific model on 1.5 million genomes, we show that GenSLMs can accurately and rapidly identify variants of concern. Thus, to our knowledge, GenSLMs represents one of the first whole genome scale foundation models which can generalize to other prediction tasks. We demonstrate scaling of GenSLMs on GPU-based supercomputers and AI-hardware accelerators utilizing 1.63 Zettaflops in training runs with a sustained performance of 121 PFLOPS in mixed precision and peak of 850 PFLOPS. We present initial scientific insights from examining GenSLMs in tracking evolutionary dynamics of SARS-CoV-2, paving the path to realizing this on large biological data.

Ruth Roberts, PhD, ATS, FBTS, ERT, FRSB, FRCPath
Director and Cofounder, ApconiX, Alderley Park UK
Chair and Director of Drug Discovery, University of Birmingham, UK

Dr Ruth Roberts is Chair and Director of Drug Discovery at Birmingham University, UK and is Cofounder of ApconiX, an integrated toxicology and ion channel company. Before that Ruth was Global Head of Regulatory Safety at AstraZeneca (2004-2014) and Director of Toxicology for Aventis in Paris, France (2002-2004). Ruth is current Chair of the HESI board of Trustees, has served on SOT council and is past president of EUROTOX, the British Toxicology Society (BTS) and of the Academy of Toxicological Sciences (ATS). Ruth was the recipient of the SOT Achievement award in 2002, the EUROTOX Bo Holmstedt Award in 2009, the SOT Founders award in 2018 and is the recipient of the 2022 ATS Millie Award, given for outstanding achievement. ApconiX recently received the 2022 Queen's Award for Enterprise. With more than 150 publications in peer-reviewed journals, Ruth is committed to developing and implementing science-led approaches to drug discovery and development.

Drug Discovery – If You Want to See Different Results Do Things Differently

Regulatory guidelines provide an important framework for the evaluation of candidate drugs to ensure patient and volunteer safety. On average, the development of a small molecule drug takes around 12 years and costs around $50m. Despite the regulatory framework and this huge investment of time and money, attrition remains a major challenge and very few molecules actually make it through to the market. Toxicities in the liver, the cardiovascular system and the CNS are prevalent and can cause drugs to stop in discovery and development. However, there are many innovative approaches that can be taken to improve success and reduce animal usage.

Early derisking of drug targets and chemistry is essential to provide drug projects with the best chance of success. Target safety assessments (TSAs) use target biology, gene and protein expression data, genetic information from humans and animals and competitor compound intelligence to understand the potential safety risks associated with modulating a drug target (1). However, there is a vast amount of information, updated on a daily basis that must be considered for each TSA.

We have developed a data science-based approach that allows acquisition of relevant evidence for an optimal TSA. This is built on expert-led conventional and artificial intelligence-based mining of literature and other bioinformatics databases. Potential safety risks are identified according to an evidence framework, adjusted to the degree of target novelty. Expert knowledge is necessary to interpret the evidence and to take account of the nuances of drug safety, the modality and the intended patient population for each TSA within each project.

Alongside understanding the potential risks associated with inhibiting or activating a drug target, it is key to evaluate the different lead candidates emerging from discovery chemistry to understand their potential for toxicity. This is frequently assessed in early 'Mini Tox' studies in the rodent and in the maximum tolerated dose/dose range finding studies (MTD/DRF) studies carried out prior to selecting one drug candidate to go forward to GLP toxicology testing. However, there is a constant drive to move away from animal testing. We have developed a deep generative adversarial network (GAN)-based framework capable of deriving new animal results from existing animal studies without additional experiments (2). Using pre-existing rat liver toxicogenomic (TGx) data from the Open Toxicogenomics Project-Genomics-Assisted Toxicity Evaluation System (Open TG-GATES), we generated Tox-GAN transcriptomic profiles with high similarity (0.997 6 0.002 in intensity and 0.740 6 0.082 in fold change) to the corresponding real gene expression profiles, proving its utility in gaining a molecular understanding of underlying toxicological mechanisms and gene expression-based biomarker development. To the best of our knowledge, the proposed Tox-GAN model is novel in its ability to generate in vivo transcriptomic profiles for different treatment conditions from chemical structures and holds great promise for generating high-quality toxicogenomic profiles without animal experimentation.

Over the past 20 years, screening for activity at cardiac ion channels such as hERG has considerably reduced attrition due to cardiovascular toxicity. Recently, we proposed that a similar approach could be taken to reduce seizure liability (3). Advances in stem cell biology coupled with an increased understanding of the role of ion channels in seizure offer an opportunity for a new paradigm in screening. We assessed the response of differentiated human induced pluripotent stem cell (hiPSC) neurones to 16 pro-seizurogenic compounds by microelectrode array (MEA). These compounds caused characteristic changes to electrical activity in key parameters indicative of seizure such as network burst frequency. Alongside the hiPSC/MEA approach, the same 16 seizurogenic compounds were screened against a panel of 15 ion channels with strong links to seizure using automated electrophysiology. We found that 15/16 compounds demonstrated at least one "hit" against the seizure panel and 9/16 compounds inhibited two or more ion channels. These studies highlight the potential utility of an integrated in vitro approach for early seizure prediction to provide mechanistic information and to support optimal drug design in early development, saving time and resources.

Overall, ion channel screening, Tox-GAN and TSAs take full advantage of the most recent developments in data science and can be used within drug projects to identify and mitigate risks, helping with informed decision making and resource management. These approaches should be used in the earliest stages of a drug project to guide decisions such as target selection, discovery chemistry options, in vitro assay choice and end points for investigative in vivo studies.

1. Roberts, RA (2018) Understanding drug targets: there's no such thing as bad news. Drug Discovery Today, 23, 1925-1928. https://doi.org/10.1016/j.drudis.2018.05.028
2. Xi Chen, Ruth Roberts, Weida Tong, Zhichao Liu, Tox-GAN: An Artificial Intelligence Approach Alternative to Animal Studies—A Case Study With Toxicogenomics, *Toxicological Sciences*, Volume 186, Issue 2, April 2022, Pages 242–259, https://doi.org/10.1093/toxsci/kfab157
3. Roberts, RA, Authier, S, Mellon, D, Morton, M, Suzuki, I, Tjalkens, RB, Valentin, J-P and Pierson, J (2021) Can we panelise seizure? Toxicological Sciences, **179**, 3-13. https://doi.org/10.1093/toxsci/kfaa167

Joseph Sanford, M.D.
Chief Clinical Informatics Officer (CCIO), University of Arkansas for Medical Sciences (UAMS)
Director, Institute for Digital Health & Innovation (IDHI)
Arkansas, USA

Joseph Sanford, M.D., is associate vice chancellor — chief clinical informatics officer (CCIO) for UAMS and the director of the Institute for Digital Health & Innovation (IDHI). He also serves as an associate professor in the College of Medicine's Department of Anesthesiology with a secondary appointment in Biomedical Informatics. He is board certified in Anesthesiology and Clinical Informatics. As CCIO, Joseph is responsible for designing and supporting the development of clinical information systems to improve the delivery of patient care throughout the UAMS Health system. His group does so with a focus on behavioral economics-influenced user experience and user interface improvements for providers and patients while maintaining, updating, and expanding the electronic medical record. The IDHI delivers real-time interactive video patient consultations, hardware and software solutions, continuing medical education, and patient education to Arkansas. IDHI has brought many programs to Arkansas including over 20 long-sustaining telemedicine initiatives, 10 distance education programs to rural providers and patients, two HRSA research centers, and technical review and monitoring of Arkansas' broadband program in partnership with the AR Dept. of Commerce. The institute represents the culmination of Arkansas' digital health expertise, with directors and stakeholders who have been instrumental in developing telemedicine initiatives in Arkansas that address the state's health disparities. Prior to assuming directorship of IDHI, Joseph was co-director, with surgeon Kevin Sexton, M.D., of the institute's Healthcare Analytics Division. They have collaborated with government, private industry, and startups to model disease spread, PPE supply and distribution, pilot new technologies, and co-develop new intellectual property via grants and other opportunities. Joseph has advised SJ Medconnect, hDrop Technologies, and Qventus, among others, and is a managing member of Datafy, LLC. An Arkansas native, Joseph received his medical degree from UAMS and his bachelor's degree with honors in computer science from the University of Arkansas. He completed a fellowship in the management of perioperative services at the Stanford University School of Medicine in Stanford, California.

Algorithmic Medicine: New Opportunities to Increase Patient Trust

Machine learning and artificial intelligence have the opportunity to change healthcare. Eric Topol, MD believes, "The greatest opportunity offered by AI is not reducing errors or workloads, or even curing cancer: it is the opportunity to restore the precious and time-honored connection and trust—the human touch—between patients and doctors." Joseph Sanford, MD and Kevin Sexton, MD will talk about the implementation of these technologies into an enterprise medical record system focusing on the challenges, opportunities, and common missteps they've encountered as Chief and Associate Chief Clinical Informatics Officers at UAMS. They'll also introduce clinical informatics, the youngest medical specialty, and what role this specialty will play in the future of these technologies.

Kevin W. Sexton, MD
Associate Professor
University of Arkansas for Medical Sciences (UAMS)
Arkansas, USA

Kevin W. Sexton, MD, is a surgeon-scientist with board certifications in Surgery and Clinical Informatics, he has used this expertise to create software that worked across multiple electronic medical records to predict patient complications in hospital and outpatient settings (Midas+ Live™) and has created medical devices designed to use venous waveforms to monitor patients. The software was acquired by Affiliated Computer Services (a Xerox company) and the device is licensed to Baxter International, both Fortune 500 companies. Kevin is currently an Associate Professor at UAMS in the Department of Surgery, with secondary appointments in the Department of Biomedical Informatics, the Fay W. Boozman College of Public Health Department of Health Policy and Management, and the UAMS College of Pharmacy Department of Pharmacy Practice, Division of Pharmaceutical Evaluation and Policy (PEP). He serves as Associate Chief Medical Informatics Officer for Innovation, Research, and Entrepreneurship at UAMS, the Associate Director of the UAMS Institute for Digital Health & Innovation, and President of BioVentures, LLC, the UAMS technology transfer office. Kevin has authored over 80 peer reviewed publications and has a track record of extramural funding from the National Institute of Health, Health Resources and Services Administration, and Department of Defense, among others. He is currently funded by the National Institute of General Medical Sciences and the National Center for Advancing Translational Sciences of the National Institutes of Health under award numbers R01 GM 111324, UL1 TR003107, KL2 TR003108, and TL1 TR003109. Technology is in Kevin's DNA. He's a managing member of Datafy and advisor to hDrop Technologies, Inc., Decisio Health, Inc., HoopCare, Inc., and others. Kevin is also a professional coach who specializes working with healthcare software as a service companies and healthcare executives.

Algorithmic Medicine: New Opportunities to Increase Patient Trust

Machine learning and artificial intelligence have the opportunity to change healthcare. Eric Topol, MD believes, "The greatest opportunity offered by AI is not reducing errors or workloads, or even curing cancer: it is the opportunity to restore the precious and time-honored connection and trust—the human touch—between patients and doctors." Joseph Sanford, MD and Kevin Sexton, MD will talk about the implementation of these technologies into an enterprise medical record system focusing on the challenges, opportunities, and common missteps they've encountered as Chief and Associate Chief Clinical Informatics Officers at UAMS. They'll also introduce clinical informatics, the youngest medical specialty, and what role this specialty will play in the future of these technologies.

Leming Shi, Ph.D.
Professor
School of Life Sciences, Human Phenome Institute, and Shanghai Cancer Center
Fudan University
Shanghai, China

Dr. Leming Shi is a professor at the School of Life Sciences, Human Phenome Institute, and Shanghai Cancer Center of Fudan University. Dr. Shi's research aims to improve the success rate of drug discovery and development and to promote precision medicine by generating and integrating high-quality multiomic data. During his tenure at the US FDA, Dr. Shi conceived and led the MicroArray and Sequencing Quality Control (MAQC/SEQC) consortium together with Dr. Weida Tong for quality control and standardization of transcriptomic and genomic data, publishing four special issues in *Nature Biotechnology*. These efforts led to the launch in 2017 of the International MAQC Society (www.maqcsociety.org) to enhance the reproducibility of high-throughput technologies. Dr. Shi served as its first president and is its Chief Science Officer and a member of the board of directors. Dr. Shi was a co-founder of Chipscreen Biosciences and co-developed the chemogenomics-based drug discovery platform, resulting in the marketing approvals of one novel HDAC inhibitor (Chidamide) for treating cancers in China and Japan, and one novel PPARα/γ/δ pan-agonist (Chiglitazar) for treating type 2 diabetes in China. More recently, Dr. Shi and his collaborators developed a multiomics-based molecular subtyping approach for precision treatment of triple-negative breast cancers, significantly improving patient survival. Dr. Shi has published over 200 peer-reviewed papers with >15,000 citations by SCI journals and an *h*-index of 60, and (co-)led the development of four ISO and CLSI standards and guidance on data quality in genomics and transcriptomics.

Quality Control and Standardization of Multiomics for Precision Medicine

Multiomics profiling is a powerful tool to characterize the same samples with complementary features orchestrating the genome, epigenome, transcriptome, proteome, and metabolome. However, the lack of ground truth hampers the objective assessment of and subsequent choice from a plethora of measurement and computational methods aiming to integrate diverse and often enigmatically incomparable omics datasets. Under the Quartet Project (http://chinese-quartet.org/), we establish and characterize the first suites of publicly available multiomics reference materials of matched DNA, RNA, proteins, and metabolites derived from immortalized cell lines from a family quartet of parents and monozygotic twin daughters, providing built-in truth defined by family relationship and the central dogma. We demonstrate that the "ratio"-based omics profiling data, i.e., by scaling the absolute feature values of a study sample relative to those of a concurrently measured universal reference sample, were inherently much more reproducible and comparable across batches, labs, platforms, and omics types, thus empower the horizontal (within-omics) and vertical (cross-omics) data integration in multiomics studies. Our study identifies "absolute" feature quantitation as the root cause of irreproducibility in multiomics measurement and data integration, and urges a paradigm shift from "absolute" to "ratio"-based multiomics profiling with universal reference materials.

Todd Shields, Ph.D.
Chancellor
Arkansas State University
Jonesboro, Arkansas, USA

Dr. Todd Shields currently serves as Chancellor of Arkansas State University. Formerly he was dean of the Fulbright College of Arts & Sciences at the University of Arkansas in Fayetteville. He joined the University of Arkansas in 1994, and has served as associate director of the J. William Fulbright Institute of International Relations, chair of the Department of Political Science, director of the Diane D. Blair Center of Southern Politics and Society, Interim dean of the Clinton School of Public Service, and dean of the Graduate School and International Education. Chancellor Shields has published dozens of journal articles and is the co-author or co-editor of several books. He has received more than $25 million dollars in grants and research support as a principal or co-principal investigator. His research focuses primarily on public opinion. He has investigated the role of public attitudes and behaviors in the areas of southern politics, elections, and public health.

A Validation Study of Fatalism Scales

Throughout most of the public health literature, Fatalism is generally defined as a “belief that the course of fate cannot be changed and that life events are beyond one’s control” and “is usually conceptualized as a set of pessimistic and negative beliefs or attitudes regarding health-seeking behaviors, screening practices, and illness” (Abraido-Lanza 2007, 153). Scholars across a range of disciplines routinely rely on measures of Fatalism as predictors of various attitudes and behaviors. For example, Fatalism is related to a range of health outcomes including reluctance to seek early screening, testing, prevention, and reactions to the Covid19 pandemic (Bayram & Shields, 2021; Blanco & Díaz, 2007; Díaz et al., 2015; Gonzales et al., 2016; Ngueutsa & Kouabenan, 2017; Shahid et al., 2020).

Despite the importance of Fatalism, there have been surprisingly few evaluations of the psychometric properties of the most widely used scales. Further, the few existing investigations are far from definitive (Valenti & Faraci 2022). Using original national survey data, we evaluate the psychometric properties of four of the most commonly used measures of Fatalism, including the Powe Fatalism Inventory, the Fatalism Scale, the FATE scale, and the Multidimensional Fatalism Scale. We assess participants’ attitudes across these measures of Fatalism and examine 1) convergent/divergent validity 2) dimensionality and 3) conduct both an exploratory and confirmatory factor analysis of each scale. The findings demonstrate which of these scales and items are most appropriate for use in the burgeoning research examining Fatalism and its influence on behavior, marketing, and public health.

Catherine Shoults, MPH
PhD Candidate, Department of Biomedical Informatics
University of Arkansas for Medical Sciences
Little Rock, Arkansas, USA

Catherine Shoults received her Bachelor of Science in biology and chemistry from Missouri State University and her Master of Public Health with an emphasis in epidemiology of microbial disease from Yale University. She is currently a PhD Candidate in the UAMS Department of Biomedical Informatics specializing in machine learning and natural language processing. Catherine started her career as a public health analyst at the Kansas Health Institute focusing on community health improvement and public health systems and services. She then worked for Quintiles (now IQVIA) as a Clinical Trial Project Manager. Prior to her time at UAMS, Catherine worked as Director of Drug Development for a small biotech company in the Kansas City metro area. She specialized in the 505(b)(2) drug pathway for small molecules using novel drug delivery manufacturing. Catherine's current research focuses on healthcare dark data and analysis of social media using Natural Language Processing. She was awarded the Arkansas Research Alliance Distinguished Performance in Artificial Intelligence Award.

Mining Reddit for Adverse Drug Reactions

There are more than 2 billion active social media users, and that number grows daily. Although social media posts contain health information regarding adverse events, the use of social media-based pharmacovigilance is in its infancy. Tapping into this under-utilized resource will support understanding the real-world effects of pharmaceutical drugs. Narrative-style social media platforms, such as Reddit, are ideal for pharmacovigilance as the user is not constrained to short posts and can discuss their health freely in a pseudo-anonymous environment. My objective is to understand the quality of Reddit as a potential source of pharmacovigilance data using Machine Learning and Natural Language Processing tools. Current pharmacovigilance suffers from gaps in who reports data and Reddit could help fill those gaps. However, the quality of Reddit social media data in the context of existing Food and Drug Administration collection tools in unknown. This talk will discuss the novel use of Reddit as a pharmacovigilance data source including the NLP techniques to mine the social media site and the quality of pharmacovigilance data found in Reddit.

Camila S. Silva, Ph.D.
Research Biologist, Division of Biochemical Toxicology
National Center for Toxicological Research (NCTR)
U.S. Food and Drug Administration (FDA)
Arkansas, USA

Dr. Camila Silva is a Research Biologist in the Division of Biochemical Toxicology at FDA's National Center for Toxicological Research (NCTR/FDA). She obtained a master's degree in immunology and parasitology from the Federal University of Uberlândia (Brazil) and a Ph.D. in biomedical research from the School of Medicine of Ribeirão Preto, University of São Paulo (Brazil). She was trained in the Division of Microbiology at NCTR/FDA as an Oak Ridge Institute for Science and Education (ORISE) postdoctoral fellow (2010-2011) to conduct research on molecular diagnosis of human coronaviruses circulating in Arkansas during the flu season. In 2014, Dr. Silva joined the Division of Biochemical Toxicology at NCTR as an ORISE postdoctoral fellow, where she has been working ever since. Dr. Silva has been involved in studies using animal models to assess the toxicity of products of interest to the FDA, including dietary supplements (Nattokinase), food contaminants (Melamine and Cyanuric Acid), among others (such as Cannabidiol and Arsenic). She investigates the use of miRNA and gene expression as molecular endpoints to assist in toxicity assessments and the molecular mechanisms by which compounds of interest to the FDA induce toxicity.

As a response to the COVID-19 pandemic, Dr. Silva used her background in virology and molecular biology to lead a project to monitor for the presence of the coronavirus SARS-CoV-2 and variants of concern in wastewater and to investigate its genomic changes during COVID-19 pandemic. Within the scope of this project, Dr. Silva established collaborations with scientists within NCTR and from the University of Arkansas for Medical Sciences and Arkansas Department of Health to maximize the impact of her study in public health.

Wastewater Surveillance and Pathogen Detection

Wastewater surveillance has been successfully used in the past to detect outbreaks of poliovirus, hepatitis A virus, norovirus, among other pathogens. It has also been implemented as an effective approach for the monitoring of SARS-CoV-2 and variants at the community level during the COVID-19 pandemic. More recently, the detection of pathogens in wastewater, such as poliovirus and monkeypox, showed promise beyond COVID-19 pandemic and emphasized the impact of the early warning of disease outbreaks on public health response. Since April 2020, we have been investigating the presence of SARS-CoV-2 and its genomic changes in wastewater influent sampled from two metropolitan areas in Arkansas. The levels of viral RNA were quantified by reverse-transcription quantitative polymerase chain reaction (RT-qPCR) targeting three different viral genes (encoding ORF1ab polyprotein, ORF1ab; surface glycoprotein, S-protein; and nucleocapsid phosphoprotein, N-protein). The identity and genetic diversity of the virus were investigated using allele-specific RT-qPCR and RNA sequencing. SARS-CoV-2 variants of concern were detected in wastewater samples throughout the duration of the study and matched those found in COVID-19 patients from Arkansas during the same period. Changes observed in the detection pattern of S- and N-genes due to viral mutations suggest that this type of data could serve as an early warning signal to investigate for new variants in the population. This study supports the use of wastewater surveillance as a reliable complementary tool for the monitoring of SARS-CoV-2 and new genomic variants at the community level, and likely other pathogens shed in the feces.

Nichole Stanley, PhD
Assistant Director of Analytics
Arkansas Center for Health Improvement (ACHI)
Arkansas, USA

Nichole Stanley, PhD, is the assistant director of analytics at the Arkansas Center for Health Improvement. Throughout her career, Dr. Stanley has worked on a variety of projects and research aimed at improving health and quality of life for Arkansans. Her most recent work includes linking de-identified medical marijuana cardholder information in the Arkansas All-Payer Claims Database (APCD) and playing a major role in the development of community-level analyses of COVID-19 in Arkansas. Dr. Stanley leads analyses and studies using the Arkansas Healthcare Transparency Initiative and the Arkansas Health Data Initiative's data warehouse. She also participates in the development of methodological approaches and enhancements to ACHI's data assets.

Kenley Money
Director of Information Systems Architecture
Arkansas Center for Health Improvement (ACHI)
Arkansas, USA

As director of information systems architecture, **Kenley Money** oversees the management of the Arkansas Center for Health Improvement's (ACHI) health information data environment. Her work is focused on leadership of the Arkansas All-Payer Claims Database solutions, supporting efforts to improve Arkansas's health payment system as it moves toward healthcare information transparency.

Money's expertise lies in health and lifestyle data integration, and she utilizes project management protocols and Information Quality principles to ensure the accuracy and timeliness of data required for health policy development and research. She manages data acquisition, validation, documentation, and training to ensure comprehensive data governance and data management for ACHI's analytical data warehouse and that all processing and data access meets HIPAA compliance requirements.

Money has over 30 years of experience in database design, management, customer data integration, and data analysis, specializing in health and lifestyle data integration. She has extensive experience in private sector marketing data management where she executed projects requiring data integration to build a single view of an individual.

Using the Arkansas Healthcare Transparency Initiative (HTI) to Estimate the Cost of Smoking with IHME's Population-Attributable Fractions

Smoking-attributable diseases and related healthcare costs affect millions of Americans. As states pursue tobacco and e-cigarette policies to offset the costs of related diseases, accurate smoking-attributable cost estimates are vital. Our study sought to provide more accurate, up-to-date, and condition-specific smoking cost estimates than were previously available to Arkansas policymakers. In this study, claims data in the Arkansas Healthcare Transparency Initiative (including the Arkansas APCD) was used to identify Medicaid and private health plan enrollees aged 30 to 64 in fiscal year 2015. Of these enrollees, we identified those with diseases that can be attributed to smoking based on research by the University of Washington. Total costs for all enrollees and those with a smoking-attributable disease were calculated for medical (outpatient and inpatient) claims. The proportion of all costs attributed to smoking was determined for the medical costs. This proportion was applied to the population's pharmacy and nursing home costs. Based on observed smoking-attributable conditions and population-specific condition-proportional costs, smoking-attributable costs were $795 million for Arkansas Medicaid and $542 million for privately insured beneficiaries, for a total impact of approximately $1.3 billion annually.

This study was shared at the 2022 National Network of Public Health Institutes conference.

Nuria Mencia Trinchant, Ph.D.
Instructor in cancer genomics, Guzman Laboratory.
Hematology/Oncology, Department of Medicine.
Weill Cornell Medicine (WCM**).**
New York, USA

Dr. Mencia Trinchant is an instructor in cancer genomics in Dr. Monica Guzman's laboratory at WCM. She obtained her PhD from the University of Barcelona in 2013 and joined Weill Cornell Medicine as a postdoc. Since she joined WCM she has focused her research on the monitoring and prevention of Acute Myeloid Leukemia (AML) using genomics and other molecular techniques. One of her goals is to develop tools to better monitor the presence of residual disease after treatment to address response and improve prognostic value of current methodologies. She developed a digital-PCR based assay to monitor AML patients bearing mutations in NPM1 gene (mutated in ~30% of AML cases). She is currently working on implementing next generation sequencing (NGS) as a tool for MRD as part of a multicenter and multidisciplinary group involving researchers from Europe and the USA. The end point of this initiative is to stablish a standardized platform, both from the wet lab and bioinformatic analytics standpoint, that defines specific targeted genes, sensitivity, and detection thresholds and that is validated across multiple institutions. Expanding the idea of identifying and tracing rare subclones involved in the progression towards disease, the lab initiated its studies in the field of clonal hematopoiesis (CH). She investigated the association between CH and the risk of transformation to hematological malignancy and they identified a preleukemic state that is detectable in the blood of healthy persons years before the onset of disease. This finding opens the possibility to intervention strategies to delay disease progression, not only in AML but also cardiovascular and other diseases. In collaboration with the Englander Institute for Precision medicine (EIPM) she developed a cost-effective deep sequencing test designed to detect CH and major germline cancer & cardiovascular risk alleles She expanded her findings by studying CH in many contexts, including astronauts before, during and after space flight. Dr. Mencia Trinchant is currently studying the dynamics and evolution of CH clones in the context of many diseases involving large cohorts that require huge sequencing efforts; a) HIV population, b) lymphoid malignancies and c) obesity and its role in the transformation to AML Dr. Mencia Trinchant has authored more than 20 papers.

Genomic Approaches Reveal Mutations in Blood Cells and their Clinical Implications

Clonal hematopoiesis (CH) is the acquisition of somatic mutations and expansion of a subset of blood cells in otherwise healthy individuals without any apparent blood parameter abnormality. CH mutations confer a fitness advantage, which leads to the expansion of a particular clone. These mutations can occur in precursor cells with self-renewal capabilities susceptible to accumulation of additional mutations and clonal evolution. Also, these mutated clones give rise to mutated immune effector cells, such as monocytes, granulocytes, and lymphocytes with a potential impact on many diseases. Therefore, the presence of CH becomes a risk factor for adverse clinical complications. Indeed, CH clones are dynamic entities that have been associated with elevated inflammation, risk of cardiovascular disease (CVD) and risk of hematological malignancy (HM) among other diseases. Studies performed using animal models have shown that the oncogenic penetrance of CH may be modulated by extrinsic factors like inflammation, microbial signals, and nutritional signals. Therefore, it is likely that the contribution of CH to disease is dependent on stochastic events and the interaction with each individual's "exposome". Since the presence of CH is a risk factor for disease (neoplasm and others), it is important to understand more about CH and what are the factors that contribute to its progression towards disease. Not only through correlative studies but also through the identification of strategies to eradicate these mutated clones to stop or prevent progression and the adverse clinical consequences derived.

Xiaowei Xu, Ph.D.
Professor, Department of Information Science
University of Arkansas at Little Rock
Arkansas, USA

Dr. Xiaowei Xu is a professor in the Department of Information Science at the University of Arkansas at Little Rock. Dr. Xu graduated from the University of Munich with a Ph.D. in Computer Science. Dr. Xu is the subject matter expert of AI and NLP for FDA. He worked as a senior research scientist and the team lead of Neural Computation at Siemens Corporate Technology for four years before joining the University of Arkansas at Little Rock. Dr. Xu has extensive research and development experience in Machine Learning, Artificial Intelligence, and Data Science. Dr. Xu has secured research grants from NSF, NIH, FDA, and industry. He has broad transdisciplinary research in Bioinformatics, Computer Vision, Natural Language Processing, and Text Mining. Dr. Xu has published over 100 peer-reviewed articles and papers. Dr. Xu is a recipient of the ACM SIGKDD Test of Time Award for his seminal work on the Density-Based Clustering Algorithm DBSCAN, which is one of the most popular clustering algorithms with 26,821 citations according to Google Scholar, and many open-source software implementations including Scikit-learn. Dr. Xu is a founder of the Mid-south Computational Biology and Bioinformatics Society.

AI for Natural Language Processing

AI for NLPs is a fast-growing area and has been widely used in a broad range of scientific disciplines. This workshop is to introduce the basic concept and application of AI for NLPs including causal inference. You will learn

1. Causal inference from free texts.
2. Language models.
3. Learning representations.
4. Integration of language models with machine learning.
5. AI and NLP applications to drug safety and pharmacovigilance.

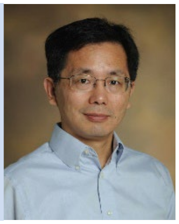
Fenghuang (Frank) Zhan, M.D. & Ph.D.
Research Director of Myeloma Center
Professor of Medicine
Morrison Family Endowed Chair in Myeloma Research
Department of Internal Medicine Slot# 508
University of Arkansas for Medical Sciences (UAMS)
Little Rock, Arkansas, USA

Dr. Fenghuang (Frank) Zhan is the Research Director of Myeloma Center at UAMS. His main expertise lies in the area of genomics and molecular/cellular biology, as well as mouse models of multiple myeloma (MM). His group has a long history of identifying treatment approaches to overcome drug resistance in MM by using genomic and genetic tools from a very large database of clinical samples and mouse models. His major contributions are (1) Defining the genetic chaos in myeloma. (2) Develop novel treatments to overcome drug resistance in multiple myeloma. (3) Understanding of tumor kinetics and microenvironmental control of myeloma growth. (4) Using gene expression profiling and proteomics to identify high-risk disease and guide therapeutic interventions. As PI, he has received many grants from National Institutes of Health / National Cancer Institute (NIH/NCI), Department of Defense (DOD), Leukemia and Lymphoma Society, Multiple Myeloma Research Foundation, Riney Foundation, Myeloma CROWD, and others. He has published over 160 peer reviewed papers including in the prestigious journals: *NEJM*, *Science*, *Cancer Cell*, *Blood, JCI*, etc.

Toward an Improved Risk Stratification for Newly Diagnosed Multiple Myeloma

Clinical outcomes of multiple myeloma are highly variable. It is unclear why current clinical and genomic variables fail to predict outcomes more accurately. Using archived data on over 1000 newly diagnosed multiple myeloma (NDMM) receiving protocol-defined total therapy at a single center for the past 22 years, we will identify molecular and clinical features linked to disease that relapses within 3 years and disease that remains in remission for > 10 years from initiation of therapy. We will focus on using archived whole genome microarray gene expression profiling (GEP) data on CD138-selected plasma cells and whole biopsies, as well as incorporating whole genome RNA sequencing (RNASeq), single cell sequencing (scSeq), and 10x Genomic single cell multiome ATAC + gene expression, and mass spectrometry proteomics (MassSpec) on archived CD138-selected cells to 1) better understand the molecular basis of myelomagenesis and 2) capture more clinical outcome variability achieved using current criteria.

Chaoyang (Joe) Zhang, Ph.D.
School of Computing Sciences and Computer Engineering
University of Southern Mississippi
Mississippi, USA



Dr. Joe Zhang is a professor in the School of Computing Sciences and Computer Engineering at the University of Southern Mississippi (USM). He was the former school director from 2008 to 20014. He established the Data Mining and Bioinformatics Lab at USM in 2003. He has maintained an active program in the areas of artificial intelligence, data mining and machine learning, and in their application domains of computational biology, bioinformatics, health informatics and toxicity analysis. He was the co-founder and steering committee co-chair of the first annual conference on Bioinformatics and Computational Biology (ACM-BCB) of the Association for Computing Machinery in 2010 and also served as President of the Midsouth Computational Biology and Bioinformatics Society (MCBIOS) in 2014-2015. His research has been supported by several federal funding agencies including the National Science Foundation (NSF), Department of Defense (DOD) and National Academy of Sciences (NAS). These funded projects include 1) Biological Network Modeling and Analysis (NSF), 2) A Novel Bayesian Learning and Optimization Approach to the Reconstruction of Gene Regulatory Networks (DOD/ERDC), 3) Novel Machine Learning Approaches to Rapid and accurate In-Silicon Toxicity Prediction of Environmental Contaminants (DOC/ERDC) and 4) Deep learning based methods for multi-omics imputation (NIH). One of his current research efforts is to develop advanced deep learning models for cancer detection using histopathological images. Dr. Zhang has graduated 15 Ph.D. students and published more than 100 papers and book chapters.

Deep Learning Based Analysis of Histopathological Images of Breast Cancer

Breast cancer is now the most commonly diagnosed cancer in the world and the disease is the leading cause of cancer mortality in women worldwide. The automatic diagnosis of breast cancer by analyzing histopathological images plays a significant role for patients and their prognosis. However, traditional feature extraction methods can only extract some low-level features of images, and prior knowledge is necessary to select useful features, which can be greatly affected by humans. Deep learning based methods can extract high-level abstract features from images automatically. In this talk, I will introduce deep learning methods for analysis of histopathological images of breast cancer and address the following questions:

- What are the impacts of image data property and preprocessing?
- How to select or develop deep learning models and perform fine-tuning?
- What metrics are used to evaluate the performance?
- How to combine machine learning techniques for performance improvement?
- What are the generalizability, challenges and promise of deep learning based methods for diagnosis of other types of cancer?

Poster Abstracts

(In alphabetical order by presenter's last name)

Poster Number: 1

Ordering-Disordering Analysis of GeSn Films using Raman Spectroscopy

Kennedy C. Abanihe, Joel Ruzindana, Wisdom Ariagbofo, Manoj K. Shah*, and Mansour Mortazavi

Department of Chemistry and Physics, University of Arkansas at Pine Bluff, Pine Bluff, AR, USA

* Corresponding author

Background: To explore the GeSn film material quality we studied the atomistic configuration through an investigation of the order-disorder analysis via Raman spectroscopy.

Method: The temperature-dependent Raman measurement from $Ge_{0.95}Sn_{0.5}$ to $Ge_{0.831}Sn_{0.169}$ was performed over a temperature range of 90K to 450K using 785nm and 532nm lasers. The photon energy of the laser-line was kept larger than the bandgap energy of GeSn films. A conventional Raman system, LabRAM-HR was used for the measurement. The Raman measurement was started at a low 90K then the temperature was increased gradually at the rate of 3K per minute and spectra were measured at an interval of 30K. The sample temperature was maintained for 3 min to ensure stable temperature during the measurement. A weak laser power was used to avoid the local heating caused by the laser and a 50x long working distance lens was used for the measurement.

Results: The measured spectra were fitted for the Ge-Ge order, Ge-Ge disorder, Ge-Sn, α-Sn, and β-Sn modes. The main Ge-Ge peak shifts left from 300 cm^{-1} showing the incorporation of Sn in the Ge lattice. The 785nm laser Ge-Sn and Sn-Sn peaks are unclear due to their high penetration depth. The 785nm laser resonance is better with Ge-Sn and Sn-Sn peaks. For the 532nm laser line Ge-Sn and Sn-Sn peaks are not suppressed and show a clear shift with Sn incorporation. The shift induced by temperature is larger than the Sn incorporation, which is mainly attributed to phonon-phonon coupling and thermal expansion. The intensity of the lower concentration Sn is comparatively lower than the higher concentration Sn.

Conclusions: The main Ge-Ge peak intensity decreases, linewidth increases, and other peaks are not clear for the samples above and below room temperature.

Keywords: Raman Spectroscopy, Phonon-phonon coupling, Ge-Ge peak intensity, LabRAM HR.

Poster Number: 2

Analyzing the Association between Aspirin and Platelet Indices in Cardiovascular Mortality Using a Decision Tree Approach

Saly Abouelenein[1], Robert Delongchamp[2], Malek Al-hawwas[3], Kevin W. Sexton[4], Carolyn J. Greene[5], Fred W. Prior[1]*

1. Biomedical Informatics, University of Arkansas for Medical Sciences, Little Rock, Arkansas, USA
2. Epidemiology, University of Arkansas for Medical Sciences, Little Rock, Arkansas, USA
3. Internal Med Cardiology, University of Arkansas for Medical Sciences, Little Rock, Arkansas, USA
4. Surgery Trauma Surgery, University of Arkansas for Medical Sciences, Little Rock, Arkansas, USA
5. Translational Research Institute, University of Arkansas for Medical Sciences, Little Rock, Arkansas, USA

* Corresponding author

Background: Cardiovascular diseases are one of the leading causes of death worldwide and early detection of those at risk of developing these diseases is crucial. One common predictor of cardiovascular mortality is platelet indices, which include measures such as platelet count, mean platelet volume, plateletcrit, and platelet distribution width. The use of aspirin, a commonly prescribed antiplatelet drug, has been shown to reduce the risk of cardiovascular events. In this study, we aim to explore the association between cardiovascular mortality, platelet indices, and aspirin use in the US population using a decision tree approach.

Methods: The National Health and Nutrition Examination Survey (NHANES) was conducted by the National Center for Health Statistics, collecting data on the health and nutritional status of the US population. Cross-sectional data from 2003 to 2018 was used to develop and validate risk prediction models, linked to the National Death Index (NDI) for information on CVD death.

Results: In our analysis, we used data on 11977 participants (5813 men and 6164 women) aged 18 years and older, randomly divided into training and blind-test groups. The decision tree identified ten subgroups with different risk patterns based on factors such as age, Body Mass Index, smoking, alcohol use, blood pressure, plateletcrit, platelet count, mean platelet volume, and aspirin use in low dose. The decision tree model achieved an accuracy of 85.5%.

Conclusion: The study was able to achieve an accuracy rate of 85.5% in classifying the effectiveness of aspirin use concerning platelet indices and its relation to cardiovascular mortality. These findings will provide essential insights into the risk factors for cardiovascular mortality and assist in developing effective strategies to prevent cardiovascular diseases.

Keywords: Cardiovascular mortality, Platelet indices, Aspirin, NHANES, Decision Tree.

Poster Number: 3

A Spatial and Spatiotemporal Analysis of Colon-Rectum Cancer in Arkansas

Anthony Adkins[1], Melody Greer[2], Sudeepa Bhattacharyya[3]

1. Department of Mathematics & Statistics, Arkansas State University, Jonesboro, Arkansas, United States
2. Department of Biomedical Informatics, University of Arkansas for Medical Sciences, Little Rock, Arkansas, United States
3. Arkansas Biosciences Institute and Department of Biological Sciences, Arkansas State University, Jonesboro, Arkansas, United States

Background: Arkansas experiences a disproportionately high age-adjusted incidence rate in Colorectal cancer (CRC) relative to the rest of the nation. In this study, geospatial and spatiotemporal analyses were used at county level to better understand the distribution of cancer incidence rates and changes in incidence rates over time in Arkansas.

Methods: Global and Local Moran's I tests were utilized to identify statistically significant age-adjusted CRC incidence rates with respect to geographic region, whereas Integrated Nested Laplace Approximation and Markov Chain Monte Carlo methods were used for Bayesian inference to discern statistical significance with a consideration to geospatial and temporal factors.

Results: The Global Moran's I test indicates that the spatial distribution of colorectal cancer incidence contains more clustering and dispersion than would be expected if the underlying spatial processes were random. The Local Moran's I test identified hotspots and clusters with respect to region, with the most concentrated region being that of counties in Northeast Arkansas. Spatiotemporal modeling shows a significant increase in CRC incidence rates in several counties in recent years.

Conclusions: The results of the study will serve to identify high-risk counties with implications that may be utilized to take concerted efforts in providing prophylactic measures to reduce the colorectal cancer incidence rate.

Keywords: geospatial analysis, spatiotemporal analysis, GIS, colon-rectum cancer

Poster Number: 4

Modeling Alzheimer-Associated Amyloid B-Peptide Interactions with P-Glycoprotein

Joseph Asante Jr.,[1] Meenakshisundaram Balasubramaniam,[2] and Steven W. Barger[2]

1. UALR/UAMS Joint Graduate Program in Bioinformatics, Little Rock, AR. USA.
2. Department of Geriatrics, UAMS, Little Rock, AR. USA.

Background: A preponderance of data indicates that Alzheimer’s disease is a consequence of excess accumulation of amyloid β-peptide (Aβ) in the brain. In rare instances, this appears to result from overproduction of Aβ, but inadequate removal of the peptide from the brain is implicated in more than 90% of cases. Prior work has suggested that the multidrug resistance protein (MDR1)—also known as p-glycoprotein (P-gp), an ATP-binding cassette (ABC) transporter—makes a major contribution to the clearance of Aβ into the blood.

Methods: In this study, we employed computational modeling and molecular dynamic simulations to examine the dynamic interactions of the two major Aβ sequences with P-gp protein. Our long-term goal is to examine the impact of P-gp substrates on the extrusion of Aβ from the brain and the impact that other P-gp substrates may have on this process.

Results: Our study suggests that both $A\beta_{40}$ and $A\beta_{42}$ monomers binds directly to P-gp and forms a stable complex during a 50-ns simulation. The trajectory of the simulation analysis of both $A\beta_{40}$-P-gp and $A\beta_{42}$-P-gp complexes shows that $A\beta_{40/42}$ is attracted to the inner chamber of P-gp protein suggesting the uptake mechanism.

Conclusion: Understanding these interactions may help in the development of health policy and therapeutic strategies for preventing Aβ accumulation and, thus, Alzheimer’s disease.

Keywords: Alzheimer's disease, Amyloid β-peptide, P-glycoprotein, Molecular modeling, and simulation.

Poster Number: 5

Addressing Underreporting Bias in Neonatal Drug Screening Rates

Johnna Berryhill[1,3], Enrique Gomez[4], Sudeepa Bhattacharyya[2,3]

1. Department of Mathematics and Statistics, Arkansas State University, Jonesboro, Ar USA
2. Department of Biological Sciences, Arkansas State University, Jonesboro, Ar USA
3. Arkansas Biosciences Institute, Jonesboro, Ar USA
4. St. Bernards Medical Center, Jonesboro, Ar USA

Background: Meconium, the first stool passed by newborns, is often screened for drugs to assess fetal exposure during pregnancy. Screening rates in this study refer to the proportion of newborns in the population of infants born at St. Bernards Medical Center from Jan. 2018-Oct. 2022 who were selected for a meconium drug screening test. Fluctuations in this screening rate can be due to a change in the method of selecting infants to be screened. This can result in biased estimates of disease prevalence.

Methods: We employed a statistical method that accounted for the unequal screening rates and allowed for the adjustment of the regression analysis to give an unbiased estimate. The results are compared to those obtained using unadjusted counts. We further analyzed the change over time in positive screening rates for each of the drugs which were screened for.

Results: The findings show that the proposed method improves the accuracy of disease prevalence estimates and reduces the impact of random fluctuations in screening rates over time. The adjusted r-squared value increased from approximately 40% (0.40, 95% CI (0.21,0.59)) to approximately 70% (0.70, 95% CI (0.56,0.81)) when using the adjusted expected counts versus unadjusted counts. Using expected counts adjusted for screening rate increases the amount of variance in meconium drug screening results explained by time by 30%. There appears to be a positive linear relationship between time and adjusted expected count with a rate of change of 0.01 per month. Of all the drugs tested, THC was the most significant to the overall increase in positive screenings.

Conclusions: This study highlights the importance of adjusting for changes in screening rates in time series data for unbiased disease prevalence estimates.

Keywords: underreporting bias, time series analysis, meconium drug screening, disease prevalence, statistical method

Poster Number: 6

SPARTAN: Self-supervised Spatiotemporal Transformers Approach to Group Action Recognition

Naga VS Raviteja Chappa[1*], Pha Nguyen[1], Alexander H. Nelson[1], Han-Seok Seo[2], Xin Li[3], Khoa Luu[1]

1. Department of CSCE, University of Arkansas, Fayetteville, Arkansas USA
2. Department of Food Science, University of Arkansas, Fayetteville, Arkansas USA
3. Department of CSEE, West Virginia University, Morgantown, West Virginia USA

[*]Corresponding author

Background: Given a video, we create local and global spatio-temporal views with varying spatial patch sizes and frame rates. We intend to classify the overall group activity of the scene by using a self-supervised framework.

Methods: The proposed self-supervised objective aims to match the features of these contrasting views representing the same video to be consistent with the variations in spatiotemporal domains. To the best of our knowledge, the proposed mechanism is one of the first works to alleviate the weakly supervised setting of GAR using the encoders in video transformers. Furthermore, using the advantage of transformer models, our proposed approach supports long-term relationship modeling along spatio-temporal dimensions.

Results: The proposed SPARTAN approach performs well on two group activity recognition benchmarks, including NBA and Volleyball datasets, by surpassing the state-of-the-art results by a significant margin.

Conclusions: We propose a new, simple, but effective Self-supervised Spatio-temporal Transformers (SPARTAN) approach to Group Activity Recognition (GAR) using unlabeled video data.

Keywords: Group Activity recognition, Self-supervised learning, Knowledge distillation

Poster Number: 7

AnimalGAN: A Generative AI Alternative to Animal Clinical Pathology Testing

Xi Chen[1], Scott Auerbach[2], Ruth Roberts[3,4], Zhichao Liu[1,5*], Weida Tong[1*]

1. National Center for Toxicological Research, Food and Drug Administration, Jefferson, AR, USA
2. Division of the Translational Toxicology, National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina, USA
3. ApconiX Ltd, Alderley Park, Alderley Edge, SK10 4TG, U.K.
4. University of Birmingham, Edgbaston, Birmingham, B15 2TT, U.K.
5. Currently working at Boehringer Ingelheim

*Correspondences: Zhichao Liu zhichao.liu@boehringer-ingelheim.com and Weida Tong Weida.Tong@fda.hhs.gov

Background: Clinical pathology testing of laboratory animals is crucial for chemical safety evaluation and risk assessment. As the toxicology community and regulatory agencies are moving towards replacement, reduction, and refinement (3Rs) of animal studies, and meanwhile, abundant animal data are available from the public domain, we are exploring an Artificial Intelligence (AI) approach to learn from the existing animal studies to generate the animal data without conducting animal experiments.

Methods: AnimalGAN was developed using Generative Adversarial Networks (GANs), which was able to learn from the associations between chemical exposure (the combination of chemical structure, dose, and exposure duration) and clinical pathology findings (hematological and biochemical measurements) in legacy animal study data in the Open Toxicogenomics Project-Genomics Assisted Toxicity Evaluation Systems (TG-GATEs) database to generate synthetic clinical pathology profiles consistent with a diverse population level response to chemical exposure.

Results: AnimalGAN successfully inferred hematological and biochemical measurements with high similarity (0.998±0.002) to the corresponding actual animal testing results under the same experimental design. Furthermore, the generated synthetic measurements by AnimalGAN could yield similar toxicity identification results comparable with those from animal samples, with an average concordance rate over 90%. Moreover, we challenged AnimalGAN to generate clinical pathology data for treatments reported in DrugMatrix. We found that the generated data can be used to assess toxicity as their corresponding actual animal data is used, with an average concordance rate around 71%, which is comparable with the concordance of toxicity assessment results by using TG-GATEs data and DrugMatrix data for their overlapped treatments.

Conclusions: The proposed AnimalGAN framework and its applications demonstrated the potential of utilizing advanced Artificial Intelligence (AI) approaches to produce non-animal models as alternatives to animal studies based on the existing data.

Keywords: Artificial Intelligence (AI); Generative Adversarial Network (GAN); Alternatives to Animal Studies; New Approach **Methodo**logies (NAMs); Clinical Pathology

Poster Number: 8

Development of a Large List of Drugs for The Study of Nephrotoxicity in Drug Discovery

Skylar Connor[1], Yanyan Qu[1], Ruth Roberts[2,3], Weida Tong[1]

1. National Center for Toxicological Research, Food and Drug Administration, Jefferson, AR, 72079, USA.
2. ApconiX Ltd, Macclesfield, United Kingdom
3. University of Birmingham, Edgbaston, Birmingham, B15 2TT, UK

Background: Drug-induced kidney toxicity (DIKI) can lead to the development of acute kidney injury, chronic kidney disease, or end-stage renal disease, causing over 1.5 million adverse events annually and affecting approximately 26% of the United States population. Currently, the standard biomarkers for DIKI identification are serum creatinine and blood urea nitrogen, both markers are late-stage biomarkers that are known to lack the sensitivity or specificity to detect nephrotoxicity prior to significant loss of renal function. Consequently, there is a pressing need for the development of alternative methods to reliably predict DIKI in early drug discovery.

Methods: For the proper development of alternative nephrotoxic methods, a large drug list with annotated DIKI potential is needed. We collected drugs from two literature datasets with confirmation using FDA drug labeling to produce a large list of drugs with known kidney toxicity, called DIKIT (Drug-Induced Kidney Injury and Toxicity). DIKIT is comprised of 1083 drugs, where 580 are DIKI positive (Nephrotoxic) and 503 are DIKI negative (non-Nephrotoxic).

Results: DIKIT covers all 14 anatomical categories with drugs related to the cardiovascular system, anti-infective for systemic use, nervous system, and alimentary tract and metabolism found to be the most prevalent. We also found that, while methods like the Rule-of-Two (RO2) and Biopharmaceutics Drug Disposition Classification System (BDDCS) are known to be successful in the evaluation and severity classification of Drug Induced Liver Injury (DILI), these methods have proven to be ineffective in the classification of a drugs nephrotoxic potential.

Conclusion: These results indicate that there are some distinct differences in nephrotoxicity as compared to DILI, such as the reabsorption, secretion, or passive filtering of drugs by the kidney. DIKIT will be a relevant and invaluable resource for the improvement of nephrotoxic research in areas such as the discovery of new methodologies to access the severity and better classify nephrotoxicity earlier within the drug development process.

Keywords: Drug-Induced Kidney Toxicity, Early Drug Discovery, In Silico Nephrotoxicity List

Poster Number: 9

Utilizing Knowledge Extraction to Establish a Quantitative Data Analysis Method on the Spectrum of Cognitive Impairment in Parkinson's Disease

Journey Eubank[1], Rohit Dhall[2], Linda Larson-Prior[3], Maryam Garza[1], John Talburt[4], Fred Prior[1]

1. Department of Biomedical Informatics, University of Arkansas for Medical Sciences, Little Rock, AR, USA
2. Department of Neurology and Neurodegenerative Disorders, University of Arkansas for Medical Sciences, Little Rock, AR, USA
3. Department of Psychiatry, Neurology, and Neurobiology & Developmental Sciences, University of Arkansas for Medical Sciences, Little Rock, AR, USA
4. Department of Information Science, University of Arkansas at Little Rock, Little Rock, AR, USA

Background: Parkinson's disease with minor to major cognitive impairment is based on a clinical diagnosis derived from a battery of clinically established screening tools. One of the most frequently used tools is the Montreal Cognitive Assessment (MoCA). The current method of use of the MoCA does not extract all the information available from the assessment but condenses patient performance into a composite score. To establish a clinical diagnosis of cognitive impairment, the MoCA score is used in conjunction with other factors such as education level, medications, and a neuropsychological evaluation. There is evidence indicating the MoCA screening tool has some clinimetric failings and that there is important information on cognition to be gleaned from patterns in the patient's performance on the exam that is being lost through the scoring protocol.

Methods: To identify these patterns, we will interview neurologists and clinical neuropsychologists specializing in Parkinson's disease and ask them to walk us through the process of how they diagnose cognitive impairment. The interviews will be guided by patient profiles containing the MoCA exams and deidentified but clinically relevant information to help the professionals elucidate patterns they look for when determining a diagnosis. We will identify the number of patterns and a list of feature descriptors based on the interviews to be used in an unsupervised algorithm. Next, we will collect and curate 500 MoCA assessments and neuropsychological assessments from the UAMS EHR where we must address data quality issues due to formatting and missing data which impacts reproducible data analysis.

Conclusions: The goal is to explore if the algorithm will reproducibly characterize the various patterns of cognitive impairment according to the interpretations of the clinical professionals.

Keywords: Knowledge engineering, Machine Learning, Parkinson's Disease, Cognitive Assessments

Poster Number: 10

Classifying Free Texts into Pre-Defined Sections Using AI in Regulatory Documents: A Case Study with Drug Labeling Documents

Magnus Gray, Joshua Xu, Weida Tong, Leihong Wu*

Division of Bioinformatics & Biostatistics, National Center for Toxicological Research, US Food and Drug Administration,
3900 NCTR Rd., Jefferson, AR 72079 USA

*Corresponding Author

Background: The US Food and Drug Administration (FDA) regulatory process often involves several reviewers who focus on sets of information related to their respective areas of review. Accordingly, manufacturers that provide submission packages to regulatory agencies are instructed to organize the contents using a structure which enables the information to be easily allocated, retrieved, and reviewed. However, this practice is not always followed correctly; as such, some documents are not well structured, with similar information spreading across different sections, hindering the efficient access and review of all the relevant data as a whole.

Methods: To improve this common situation, we evaluated an Artificial Intelligence (AI)-based Natural Language Processing (NLP) methodology, called Bidirectional Encoder Representations from Transformers (BERT), to automatically classify free-text information into standardized sections, supporting a holistic review of drug safety and efficacy. Specifically, FDA labeling documents were used in this study as a proof of concept, where the labeling section structure defined by the Physician Label Rule (PLR) was used to classify labels in the development of the model. The model was subsequently evaluated on texts from both well-structured labeling documents, (i.e., PLR-based labeling), and less- or differently structured documents, (i.e., Non-PLR and Summary of Product Characteristic [SmPC] labeling).

Results: In the training process, the model yielded 96% and 88% accuracy for binary and multi-class tasks, respectively. The testing accuracies observed for the PLR, Non-PLR, and SmPC testing datasets for the binary model were 95%, 89%, and 89%, and for the multi-class model, were 84%, 74%, and 67%, respectively.

Conclusions: Our study demonstrated that automatically classifying free texts into standardized sections with AI language models could be an advanced regulatory science approach supporting the review process by effectively processing unformatted documents.

Keywords: Drug labels; artificial intelligence; natural language processing; unstructured data processing; text classification

Disclaimer: The authors accept sole responsibility for the statements and accuracy in the abstract. This presentation is not a formal dissemination of information by the FDA and does not represent agency position or policy.

Poster Number: 11

Gonadotrope Microscopy in The Intact Murine Pituitary: A Novel Approach for Understanding Pituitary Cellular Networks

Ashley Herdman[1], Alex Lagasse[1], Kenzie MacNicol[1], Anessa Haney[1], Ulrich Boehm[2], Melanie MacNicol[1], Angus MacNicol[1], Gwen Childs[1], James Hyde[3], Angela K. Odle[1]

1. Department of Neurobiology and Developmental Sciences, University of Arkansas for Medical Sciences, Little Rock, Arkansas, United States of America
2. Department of Pharmacology and Toxicology, Saarland University, Homburg, Germany
3. Department of Biology, Southern Arkansas University, Magnolia, Arkansas, United States of America

Background: Despite the importance of pituitary hormones for development, reproduction, and metabolism, the mechanisms through which pituitary cells coordinate are still largely unknown. Ex-vivo imaging of pituitary networks has been primarily conducted through pituitary slices, drastically disrupting the structure, and potentially function, of the pituitary. This limited methodology remains an obstacle to understanding pituitary communication and the mechanisms through which endocrine signaling occurs. We have developed and validated novel methodology for imaging gonadotropes, the reproductive cells of the pituitary, without disturbing the gland, and have begun investigating the impact of external stimulation on functional network coordination.

Methods: To visualize gonadotrope calcium signaling, we created a mouse line expressing *Gnrhr*-driven Cre for gonadotrope specificity and a Cre-dependent fluorescent calcium indicator. Intact pituitaries were taken from adult female mice during proestrus, and gonadotrope network responses to vehicle and gonadotropin-releasing hormone (GnRH) at various doses (100 nM, 500 nM, or 1 uM, n ≧ 4 pituitaries) were recorded using confocal microscopy. MATLAB software was used to identify and track signaling patterns of individual gonadotropes and compare overall gonadotrope network coordination during vehicle/GnRH stimulation. Through the MATLAB toolbox “mic2net”, cross correlation matrix analyses were performed to determine metrics of overall network strength, notably connectivity.

Results: We identified, tracked, and analyzed signaling of ~354 ± 31 gonadotropes per recording, validating the efficacy of our model. When compared to vehicle stimulation, gonadotrope connectivity increased during high dose GnRH stimulation. Connectivity did not significantly increase after 100 nM stimulation but did increase by 180% during 500 nM GnRH stimulation ($p<0.001$), and 94% during 1 uM GnRH stimulation ($p<0.01$).

Conclusions: We have developed and validated novel methodology for studying pituitary networks and analyzing cell-type specific signaling. Using this methodology, GnRH was established as a driver of gonadotrope network coordination in proestrus pituitaries.

Keywords: Reproduction, pituitary, network

Poster Number: 12

VAST (Voice and Spiral Tool):

A Novel Multimodal Machine Learning Method to Detect Parkinson's Disease and Assess Severity

Anu Iyer[1], Dr. Fred Prior[2], Dr. Tuhin Virmani[2], Dr. Linda Larson-Prior[2], Dr. Yasir Rahmatallah[2], Mr. Aaron Kemp[2], Ms. Lee Conrad[1]

1. Little Rock Central High School, Little Rock, Arkansas, United States
2. Department of Biomedical Informatics, University of Arkansas for Medical Sciences (UAMS), Little Rock, Arkansas, United States

Background: Parkinson's disease (PD) is a neurodegenerative disorder prominent in individuals 65 years and older. Currently, there are challenges with detecting PD: specialists have only 80%-84% accuracy with a 15-90 minute testing period in a clinical setting, which adds a burden to the elderly population who suffer restricted mobility. In our rural state, three out of the four trained neurologists practice at one institution and 75% of our population resides in medically underserved areas. These issues underscore the need for an accurate virtual PD diagnostic tool that implements a proven feature set already incorporated in in-person clinical assessments.

Methods: After a literature review, a 32GB RAM laptop, Inception v3 Architecture, and Jupyter Notebook were obtained. Voice data were collected through UAMS voicemails as *.wav* files and converted to 500x500 *.jpg* spectrogram files. Spiral data were collected in the Virmani Gait Lab as *.pdf* files and were cropped to 500x500 *.jpg* files. Both data types were deidentified and used in separate transfer learning convolutional neural networks with 0.001 learning rate and 15 training epochs. Training and validation accuracy-loss graphs, and confusion matrices were created by a Python program, and feature explainability was conducted via Grad-CAM.

Results: Through a multimodal approach, our proposed tool demonstrates a 96% accuracy for PD diagnosis (*'Ah'* 3-second voice test: 92% accuracy for diagnosis) and assessment of severity (hand-drawn Archimedes spirals: 100% accuracy for mild or severe PD). Our Grad-CAM implementation highlights specific features between classifications that can be analyzed in-depth.

Conclusions: Project VAST is successful in providing an accurate and effective method for PD diagnosis in a clinical or virtual setting through vocal and handwriting feature-based machine learning models. VAST may ultimately aid in accelerating PD diagnosis, resulting in improved clinical outcomes.

Keywords: Parkinson, diagnosis, transfer learning, voice, spiral

Poster Number: 13

Comparative Evaluation of Long-term Consumption of High-Fat Diet on the Expression of Alzheimer's Disease Related Genes in the Ileal Mucosa of Wild Type and Alzheimer's Disease Rat Model

Kumari Karn[1], Kuppan Gokulan[1], Sumit Sarkar[2], James Raymick[2], and Sangeeta Khare[1]

1. Division of Microbiology, National Center for Toxicological Research, US Food and Drug Administration, Jefferson, AR USA
2. Division of Neurotoxicology, National Center for Toxicological Research, US Food and Drug Administration, Jefferson, AR USA

Background: A high-fat diet (HFD) is shown to have a profound effect on the gut and brain, and its consumption is linked to dementia accelerating *Alzheimer's disease (AD)* pathology.

Methods: The aim of this study is to assess the effect of HFD on the expression of AD related genes in the gastrointestinal tract (GIT), using the RT-qPCR gene expression method in both wild-type (WT) and APP/PS1 overexpressed AD rat model.

Results: Out of a total of 84 AD related genes, the metabolic consequences of HFD contributed to 52.4% of altered gene expression both WT and AD rats. Interestingly, AD related genes were expressed in higher number in the WT rats (88.6%) as compared to the AD rats (36.4%). The male rats showed significantly lower (9.09%) expression of those genes in the AD model than in WT (70.5%). Likewise, female rats also showed a similar pattern of lower gene expression in the AD model (32%) than in WT (52.3%). In female WT rats, expression of apoptotic genes such *App* (*amyloid beta precursor protein*) and *EP300* (*E1A binding protein p300*) was markedly increased due to HFD. Furthermore, in the HFD fed AD rats there was overexpression of genes like *Bdnf* (*Brain derived neurotrophic factor*) in females and *Chat (choline acetyltransferase)* in males, which are crucial for synaptic transmission.

Conclusions: The higher percent of AD related genes expressed in WT on a HFD suggests that long-term consumption of HFD may have an impact on the expression of key neurotransmitters. The findings also show that the females are relatively more vulnerable to developing AD while consuming HFD. Studies are ongoing to assess the effect of HFD on the intestinal microbiome and secretory signals, and how it impacts gut-brain axis in the WD and AD model of this diseases.

Key words: gut-brain axis; pro-inflammatory microbes; neuroinflammation; microbiome; Alzheimer's disease

Poster Number: 14

Genomic Characterization of *Enterococcus* Species Infecting Broiler Embryos

Aishat O. Lawal, Abass Oduola, Khawla Alharbi, Andi Asnayanti, Layla Almitib, Alyssa Papineau, Ruvindu Perera, Adnan Alrubaye and Douglas D. Rhoads*

Program in Cell and Molecular Biology, University of Arkansas, Fayetteville, AR 72701, USA

*Corresponding author Email: drhoads@uark.edu

Background: We are assembling genomes for isolates of *Enterococcus avium* and *Enterococcus gallinarum* infecting egg yolks or aborted embryos, from broiler flocks. This project is focused on 1) whether specific bacterial pathogens can be vertically transmitted from breeder hens to subsequent broiler flocks, and 2) the genetic comparison of human and chicken isolates of *E. avium* and *E. gallinarum*.

Method: Fertilized eggs (n=240) from commercial broiler breeders were incubated at 99.5^{o}F and 60% relative humidity with automatic rotation. On the 18th day of incubation, eggs were candled to identify developing embryos. Yolks from apparently non-fertile eggs were sampled using a sterile swab. Swabs, liver, and swollen gizzards samples were inoculated onto plates of CHROMagar Orientation. For 30 unincubated eggs the yolk was separated and cultured overnight with an equal volume of tryptic soy broth +1% chicken serum. Positive cultures for yolk or embryo were purified and species determined by sequencing of the V1-V5 region of 16S rDNA. Total DNA was purified from isolates and sequenced.

Result: One live embryo yolk sac swab produced 10 colonies of *E. gallinarum.* Yolk sac swabs from two early aborted embryos produced bacterial lawns, which were identified as *E. gallinarum* and *Enterococcus faecalis.* Two nonfertile or very early dead yolks yielded lawns of *E. gallinarum,* and *Globicatella sanguinis*. The current NCBI database contains 45 genomes for *E. avium* (29 from human, 1 from chicken, and 15 with no-host or other hosts), and 98 genomes for *E. gallinarum* (37 human/hospital, 12 chicken with the remainder from cattle, mouse, pig, cow or host not specified). None of the chicken isolates of *E. gallinarum* genomes in NCBI appear to be from a disease state.

Conclusion: The comparative analysis of *E. gallinarum* and *E. avium* would provide knowledge into their evolution of pathogenicity and host-preference and their role in egg fertility and hatchability.

Keywords: Enterococcus, embryo, vertical transmission, pathogenicity

Poster Number: 15

Data Mining of Opioid-related Adverse Events from the FDA Adverse Events Reporting System

Huyen Le[1], Huixiao Hong[1], Weigong Ge[1], Henry Francis[2], Beverly Lyn-Cook[3], Yi-Ting Hwang[4], Paul Rogers[1], Weida Tong[1], Wen Zou[1]*

1. Division of Bioinformatics and Biostatistics, 3Division of Biochemistry Toxicity, National Center for Toxicological Research, U.S. Food and Drug Administration, Jefferson, AR, USA.
2. Office of Translational Science, Center for Drug Evaluation and Research, U.S. Food and Drug Administration, Silver Spring, MD, USA.

[4]Department of Statistics, National Taipei University, Taipei, Taiwan.

Background: Opioids are effective in treating pain, they however can cause a variety of side effects. The current opioid epidemic (i.e., an increasing number of opioid use and overdose deaths) is a serious national crisis which affects public health as well as social and economic welfare in the United States. Researchers have been interested in performing an in-depth analysis of opioid-related adverse events (AEs). Currently there is a gap on comprehensive studies which provide a global view of AEs with various opioids from the post-marketing databases.

Methods and Results: A total of 14,970,399 AE reports were retrieved and downloaded from the FDA Adverse Events Reporting System (FAERS) from 2004 Quarter I to 2020 Quarter III. After the normalizations, we obtained 20,178,515 pairs of drug-AE, originated from 69,889 unique drugs and 22,260 AEs. Then, 78,874 pairs of opioid-AEs were extracted including 13 FDA-approved opioids and 14,374 unique AEs. Among these 78,874 pairs, only 3,317 pairs were identified as the potential safety signals according to the results of Empirical Bayes Geometric Mean (EBGM) and EB05 analysis, which included 13 FDA-approved opioids and 2,709 unique AEs. The top 10 most reported AEs of each of the 13 opioids were calculated and compared both on the number of the reports and percentages among the opioids. The relationships of the 13 opioids were revealed by the network analysis and hierarchical clustering analysis.

Conclusions: The results of our comparative study on AEs associated with different opioids provide a global overview of the current status of opioid-associated AEs and the potential alternatives to avoid severe AEs in opioid prescription. The results also improve the knowledge and insights for patients, physicians, and healthcare providers in terms of safe opioid usage.

Key words: prescription opioids, AEs, FAERS, data mining, potential safety signals.

Poster Number: 16

Assessments of TMB estimation by targeted panel sequencing, a comprehensive simulation analysis

Dan Li, Binsheng Gong, Dong Wang, Huixiao Hong, Weida Tong, Joshua Xu*

Division of Bioinformatics and Biostatistics, National Center for Toxicological Research, US Food and Drug Administration, Jefferson, AR 72079, USA

*Corresponding author

Background: Tumor mutational burden (TMB), when at a high level, is an emerging indicative factor of sensitivity to immune checkpoint inhibitors. Previous studies have shown that the more affordable and accurate targeted panels can be used to measure TMB as a substitute for whole exome sequencing (WES). However, additional processes, such as hotspot mutations exclusion and TMB adjustment, are usually required to deal with the effect of the limited panel sizes. A comprehensive investigation of the effective factors is needed for accurate TMB estimation by targeted panels.

Methods: In this study, we quantitatively evaluated the variances of TMB values calculated by WES and targeted panels using 10,000 simulated targeted panels with panel sizes ranging from 0.2 to 3.1 million bases. With The Cancer Genome Atlas (TCGA) cancer samples and mutation profiles, we fixed regressions on WES-TMBs and panel-TMBs to assess the performance of a given targeted panel.

Results: Our study confirms that panel size is one of the major factors impacting TMB estimation. In addition, we were able to fit a formula that describes the relationship between panel sizes and TMB estimation variations. In addition, we identified well-performing small panels that reported TMB values similar to those obtained through WES. By analyzing data from TCGA, we demonstrated the cancer type-specific impacts of genes on TMB estimation and identified high-impact gene sets for different cancer types.

Conclusions: For cancer patients who have been diagnosed using targeted panels, accurate TMB estimation using targeted panels is then helpful in determining the most appropriate treatment plan, particularly for immunotherapies. This highlights the importance of developing targeted panels that can provide reliable and precise TMB measurements. Our study's findings on the quantitative correlations between TMB variance and panel size, and the impact of individual genes on TMB estimation, offer valuable insights into the design and utility of targeted panels for TMB measurement in cancer samples.

Key words: Tumor mutational burden, TMB, targeted panel sequencing, simulation analysis

Poster Number: 17

Random Forest Model for Predicting μ Opioid Receptor Binding Activity for Assisting Development of Opioid Drugs

Zoe Li, Jie Liu, Fan Dong, and Huixiao Hong*

Division of Bioinformatics and Biostatistics, National Center for Toxicological Research, U.S. Food and Drug Administration, 3900 NCTR Road, Jefferson, AR 72079, USA

*Corresponding author

Background: The opioid epidemic is one of the most prominent and severe public health crises in U.S. history. The devastating consequences of the opioid crisis are not only the increased number of deaths caused by opioids but also the increased economic burden of combating this crisis. The highly addictive nature of the opioids is closely related to the overdose fatalities caused by prescription opioids, heroin, and illicit fentanyl. However, the therapeutic benefits of prescription opioids acting as the most potent analgesic make prohibition of the drug impossible. Opioids exert its analgesic effect by binding to the μ opioid receptor (MOR), which then activates its downstream signaling pathway, eventually leading to the inhibition of spinal cord pain transmission. Since the discovery of MOR in the 1970s, many efforts have been endeavored to understand the structure activity relationship between the receptor and its ligands, hoping to shed some lights on the development of non- or less-addictive opioid analgesics. Yet, many questions remain unanswered, and the development of non- or less-addictive opioid analgesics has had limited success.

Methods: To understand the structure activity relationship between the MOR and its ligands, we developed a machine learning model that can be used to predict the binding activity of small molecule compounds to the MOR based on chemical structures. We first curated 17,856 MOR binding data points for 11,427 chemicals from diverse data sources such as publicly available databases, patents, and the literature. Mold2 descriptors were calculated for these chemicals. The 11,427 chemicals were then split into a training set of 5,729 chemicals with even identification numbers and a testing set of 5,698 chemicals with odd identification numbers. Random forest algorithm was employed for predictive model development. The random forest models were evaluated using 500 runs of 5-fold cross validations on the training data set and were challenged with the testing data set.

Results: Our result have shown 91.4% and 90.9% MOR binding activity prediction accuracy in the cross validations and external validation, respectively.

Conclusions: In summary, our proposed model for predicting binding activity of small molecules to opioid receptor may be useful to identify novel MOR binders, which may aid the development of new drugs targeting on MOR.

Keywords: opioid receptor, binding, agonist, antagonist, machine learning

Disclaimer: This abstract reflects the views of the authors and does not necessarily reflect those of the U.S. Food and Drug Administration.

Poster Number: 18

Analyzing Methylation Patterns of Cis-regulatory Regions

Li Ma[1,2], Erich A. Peterson[1], Donald J. Johann, Jr.[1*]

1. Winthrop P. Rockefeller Cancer Institute, University of Arkansas for Medical Sciences, Little Rock, AR, United States
2. Department of Information Science, University of Arkansas at Little Rock, Little Rock, AR, United States

[*]Corresponding author

Background: Cancer begins as a genomic disease at the cellular level, and this commonly involves mutations and altered epigenetic programming. Cancer specific changes in DNA methylation are recognized and often are focused on hypermethylation of normally unmethylated CpG islands of promoter and enhancer regions. For instance, gene silencing may lead to an accelerated carcinogenesis, especially if the gene is involved with tumor suppression or DNA repair functions and, are outcomes of promoter hypermethylation. Enhanced interrogation of vital genomic regions (eg, promoter, enhancer) are needed. To address this problem, our group is developing a software framework that can visualize and report the methylation patterns of cis-regulatory elements of a specified gene.

Methods: Tumor and normal tissues were collected from a patient diagnosed with a rare thoracic neoplasm. An enzymatic-based methylation sample prep was performed, having a targeted panel covering more than four million genomic regions, for Next Generation Sequencing (NGS). To ensure reproducibility and accuracy, the NGS post-pipeline accuracy and reproducibility system (NPARS) was utilized, and enhanced by incorporating the cis-regulatory analysis framework.

Results: Methylation status and the patterns of cis-regulatory regions of a given gene are performed using visual plots and comprehensive reports for tumor and normal tissue specimens. Visualizations include overlaid annotations of CpG islands integrated with: transcription start sites (TSS) and promoter regions, the gene body, known enhancer regions.

Conclusions: DNA methylation plays an import role in carcinogenesis. There is an unmet need for improved tools for visualizing and reporting methylation patterns within NGS datasets. To this end, we are developing a new analytical approach. This new framework is being integrated into the NPARS system for visualizing and reporting methylation patterns of cis-regulatory annotated genomic regions in an accurate and reproducible manner.

Keywords: DNA methylation, cis-regulatory, cancer, reproducibility, epigenetic

Poster Number: 19

The Effects of Disease Prevalence on Predictive Data Analysis Metrics

Thomas McCall*, Paul Rogers, Dong Wang

Division of Bioinformatics and Biostatistics, National Center for Toxicological Research, U.S. Food & Drug Administration, Jefferson, AR 72079, USA

* Correspondence: Thomas.McCall@fda.hhs.gov

Background: Screening tests for disease have their performance measured through sensitivity and specificity, which inform how well the test can discriminate between those with and without the condition. Typically, high values for sensitivity and specificity are desired. These two measures of performance are unaffected by the outcome prevalence of the disease in the population. Research projects into the health of the American Indian frequently develop Machine learning algorithms as predictors of conditions in this population. In essence, these models serve as in-silico screening tests for disease.

Methods: A diagnostic test's sensitivity and specificity values, typically determined during the development of the test, inform on the performance at the population level and are not affected by the prevalence of disease. The positive predictive value (PPV) of a screening test are susceptible to the outcome prevalence. As the number of artificial intelligence and machine learning models flourish to predict disease outcomes, it is crucial to understand if the PPV values for these in-silico methods suffer as traditional screening tests in a low prevalence outcome environment. The Strong Heart Study (SHS) is an epidemiological study of the American Indian and has been utilized in predictive models for health outcomes. We used data from the SHS focusing on the samples taken during Phases V and VI.

Results: Logistic Regression, Artificial Neural Network, and Random Forrest were utilized as in-silico screening tests within the SHS group. Their sensitivity, specificity, and PPV performance were assessed with health outcomes of varying prevalence within the SHS subjects. Although sensitivity and specificity remained high in these in-silico screening tests, the PPVs' values declined as the outcome's prevalence became rare.

Conclusion: Machine learning models used as in-silico screening tests are subject to the same drawbacks as traditional screening tests when the outcome to be predicted is of low prevalence.

Keywords: Artificial Intelligence, Machine Learning, Screening Test, Prevalence, Rare.

Poster Number: 20

Cancer Variant Allele Sequencing (CANVAS): A Computational Pipeline for The Quantitation of Ultra-Low Frequency Hotspot Mutations in Myeloid Neoplasm-Associated Genes Via Targeted Error-Corrected Sequencing of Human Peripheral Blood DNA

Page B. McKinzie*, Jennifer B. Faske, Lascelles E. Lyn-Cook, Jr., and Meagan B. Myers*

Division of Genetic and Molecular Toxicology, National Center for Toxicological Research, US Food and Drug Administration, Jefferson, AR, USA

*Corresponding authors

Background: Error-corrected sequencing describes a set of sequencing protocols focused on mitigating the effects of induced errors in calling variants and often used for evaluating somatic mutations associated with carcinogenesis. Several laboratory techniques can be used but require a matching data processing computational workflow to process the specialized raw sequencing data to accurately quantify variants of somatic origin.

Methods: Here we describe the preparation of uniquely barcoded DNA libraries of synthesized sequences in known amounts and previously evaluated genomic DNA (gDNA) and the CANVAS workflow based on mainly standard sequencing software tools to output sequences corrected for errors introduced during PCR, library preparation and sequencing. Mutant fraction samples of 10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}, and 10^{-5} were generated using synthesized gene blocks containing a set of known blood-based cancer mutations while gDNA was evaluated using samples previously evaluated with ACB-PCR and ddPCR. Performance of CANVAS was evaluated using both mutant fraction standards and comparing gDNA to previous ACB-PCR and ddPCR results (ranging from 10^{-1} to 10^{-5}). Target amplification adds molecular tags to both ends of amplicons and the CANVAS program uses these to provide accurate counts of sequences. The logic of CANVAS removes the introduction of unidentifiable nucleotides (“N”) by comparing sequences with the same UID as whole words rather than single letters, making it less noisy and using less computational time.

Results: The results show this method accurately evaluates the abundance of mutations at each position over the range of 10^{-1} to 10^{-5} as determined using the constructed mutant fraction samples and comparison to results obtained from orthogonal methods.

Conclusions: Application to future avenues of research include iPSC-derived biologic product safety assessment, and blood-based biomarkers of risk for cancer (*e.g*. therapy-related or after Phase I clinical trials) and noncancer diseases of aging and inflammation.

Key words: error-corrected sequencing, mutation, bioinformatics

Poster Number: 21

Comparative Pharmacokinetics of Zileuton's Active Pharmaceutical Ingredient, Nanocrystal-Drug Formulation, and Physical Mixture in Male and Female Sprague Dawley Rats

Chandra Mohan Reddy Muthumula[1], Bhagya Wickramaratne[1], Sangeeta Khare[1], Sushanta Chakder[2], and Kuppan Gokulan[1]

1. Division of Microbiology, National Center for Toxicological Research, US Food and Drug Administration, 3900 NCTR Road, Jefferson, AR, 72079
2. Center for Drug Evaluation Research, US Food and Drug Administration, White oak, Silver Spring, MD, 20993

Background: The biopharmaceutical classification system (BCS) is used to determine the bioavailability of an oral drug based on its solubility and permeability properties. Zileuton, a 5-lipoxygenase inhibitor used to treat chronic asthma, is classified as BCS class II drug due to its poor solubility and high permeability. To address solubility issues and improve bioavailability of a BCS class II drug, Zileuton was used as a model drug to develop a nanocrystal-drug (ND) formulation.

Methods: The pharmacokinetics of Zileuton were evaluated in 10-week-old Sprague Dawley rats following oral administration of active pharmaceutical ingredient (API 30 mg/kg bw); ND and physical mixture (PM) (30 mg each containing 7.5 mg of API) in gelatin capsules. HPLC was used to measure zileuton levels in the ileum (15 days), plasma (1, 2, 4, 6, and 24h), and urine (24h) post-treatment samples.

Results: The results of this study showed that female rats had lower Zileuton concentrations in the ileum after oral administration of the API and ND formulation than male rats. Female rats had higher plasma Zileuton concentrations after API, ND, and PM treatment than male rats. The lower plasma Zileuton concentration in male rats correlated with higher expression of phase-1 and phase-II metabolic enzymes in intestinal tissue. The Tmax of Zileuton in the male rats was found to be at 2h after ND and PM treatment, compared to 1h after API treatment. In the female rats, the Tmax of Zileuton was 1h when treated with ND and PM, but 2h when treated with API. Following ND and PM treatments, male rats had higher zileuton urine concentrations than female rats.

Conclusions: Pharmacokinetic variations of Zileuton were observed in both male and female rats following oral administration of API, ND, and PM. This study justifies further investigation of the ND in the translational model of asthma.

Poster Number: 22

UTOPIA: Unconstrained Tracking Objects without Preliminary Examination via Self-Supervised Cross-Domain Knowledge Transfer

Pha Nguyen[1*], Kha Gia Quach[2], John Gauch[1], Khoa Luu[1]

1. Department of Computer Science and Computer Engineering, University of Arkansas, Fayetteville, Arkansas, USA
2. pdActive Inc., Longmont, Colorado, USA

*Corresponding author

Background: The goal of Multiple Object Tracking (MOT) is to identify the bounding boxes and identities of targeted objects in consecutive video frames. Although fully-supervised MOT methods have performed well on existing datasets, they are not capable of adapting to new data or unseen domains.

Methods: In this study, we tackle the MOT problem from a cross-domain perspective by imitating the process of acquiring new data in practice. A novel self-supervised cross-domain MOT adaptation method is introduced that can learn and improve from target data feedback without relying on prior human knowledge of object understanding and modeling.

Results: We conduct extensive experiments on challenging settings including MOTSynth to MOT17 and MOT17 to VisDrone and demonstrate the adaptability of our proposed self-supervised learning strategy. The results also show that our method outperforms state-of-the-art methods in terms of tracking metrics MOTA and IDF1, including fully-supervised, unsupervised, and self-supervised methods.

Conclusions: This work presents the MOT problem from the cross-domain viewpoint. Furthermore, it propose a new MOT domain adaptation without any pre-defined human knowledge in understanding and modeling objects. Still, it can learn and update itself from the target data feedback. Through intensive experiments on two settings, we first prove the adaptability on self-supervised configurations and then show superior performance on tracking metrics MOTA and IDF1, compared to fully-supervised, unsupervised, and self-supervised methods.

Keywords: Multiple Object Tracking, Self-Supervised Learning, Domain Adaptation

Poster Number: 23

Fairness in Visual Clustering: A Novel Transformer Clustering Approach

Xuan-Bac Nguyen[1*], Chi Nhan Duong[2], Marios Savvides[3], Khoa Luu[1]

1. Department of Computer Science and Computer Engineering, University of Arkansas, Fayetteville, Arkansas, USA
2. Concordia University, Canada
3. Department of Electrical and Computer Engineering, Carnegie Mellon University, USA

*Corresponding author

Background: Promoting fairness for deep clustering models in unsupervised clustering settings to reduce demographic bias is a challenging goal. This is because of the limitation of large-scale balanced data with well-annotated labels for sensitive or protected attributes.

Methods: In this paper, we first evaluate demographic bias in deep clustering models from the perspective of cluster purity, which is measured by the ratio of positive samples within a cluster to their correlation degree. This measurement is adopted as an indication of demographic bias. Then, a novel loss function is introduced to encourage a purity consistency for all clusters to maintain the fairness aspect of the learned clustering model. Moreover, we present a novel attention mechanism, Cross-attention, to measure correlations between multiple clusters, strengthening faraway positive samples and improving the purity of clusters during the learning process.

Results: Experimental results on a large-scale dataset with numerous attribute settings have demonstrated the effectiveness of the proposed approach on both clustering accuracy and fairness enhancement on several sensitive attributes.

Conclusions: To the best of our knowledge, our work is one of the first studies to address the fairness issue in large-scale visual clustering. Our framework contributes not only to fairness but also to clustering performance overall.

Keywords: Fairness, Unsupervised Learning, Clustering, Transformers, Deep Learning.

Poster Number: 24

Transcriptomic And Molecular Modeling Analysis of The Effects of The Vitamin E Related Compounds in Mitochondrial Function

[1]E. Nathalie Pineda, [1]Shivangi Shrimali, [1]Ujwani Nukala, [1]Awantica Singh, [1,3]Stephen Shrum, [2]Shraddha Thakkar, [1]Philip Breen, [1]Rupak Pathak, [2] Weida Tong and [1]Cesar M. Compadre

1. Department of Pharmaceutical Sciences, University of Arkansas for Medical Sciences, Little Rock, AR 72205, USA
2. Division of Bioinformatics and Biostatistics, National Center for Toxicological Research, US Food and Drug Administration, Jefferson, AR 72079, USA
3. Tocol Pharmaceuticals, LLC, Little Rock, AR 72205, USA

Background: Vitamin E components are strongly protective against radiation exposure. Vitamin E is comprised of eight tocols, which are classified as tocopherols or tocotrienols depending on the side chain saturation and the methylation pattern on the chromanol head. Generally, the therapeutic effects of vitamin E tocols are associated with their antioxidant properties, which are key to their radioprotective effects. In cells, mitochondria produce reactive oxygen species as a byproduct of metabolism, however mitochondria are also very sensitive to oxidative damage. Although the mitochondrial-protective effects of vitamin E are recognized, the mechanisms underlying it are not well understood. In this research we have analyzed the genomic effects on mitochondrial function for a set of tocols on endothelial cells (HUVEC) exposed to radiation.

Methods: Transcriptomic changes were profiled and compared between unirradiated cells and cells exposed to 2.5 Gy radiation with pretreatment of either vehicle or therapeutic concentrations of alpha-, gamma-, or delta-tocotrienol. Ingenuity pathway analysis (IPA) analyzed the transcriptomic data and predicted the effects on oxidative phosphorylation and synthesis of ATP.

Results and conclusions: Radiation alone decreased ATP synthesis by inhibiting Complexes III and V and downregulating cytochrome c. All three tocols (AT3, GT3, and DT3) reversed the effect of radiation and upregulated ATP synthesis; however, they acted on different electron transport chain (ETC) complexes. AT3 upregulated Complex I and V, but had no effect on Complex III or cytochrome c. DT3 and GT3 upregulated Complex III, IV, V, and cytochrome C, but had no effect on Complex I. These tocols may be upregulating oxidative phosphorylation during radiation injury through different sites on the ETC. Molecular modeling analysis of inhibition of Complex III by tocols, at supraphysiological concentrations, shows that the interaction of these compounds may be modulated by the number of methyl groups on the chromanol head and by the overall lipophilicity of the molecules.

Keywords: Vitamin E, mitochondria, transcriptomic analysis, molecular modelling.

Poster Number: 25

DICTrank: The Largest Drug-induced Cardiotoxicity Reference List Annotated Based on FDA-Approved Drug Labeling Documents

Yanyan Qu[1,2], Dongying Li[1], Ting Li[1], Zhichao Liu[1], Weida Tong[1]*

1. National Center for Toxicological Research, US Food and Drug Administration, Jefferson, AR, USA 72079
2. University of Arkansas at Little Rock and University of Arkansas for Medical Sciences Joint Bioinformatics Program, Little Rock, AR, USA 72204

Background: Drug-induced cardiotoxicity (DICT), toxicity to the heart caused by medications, is one of the leading causes for drug failure in clinical trials or withdrawal from the market. As a result, many studies have been proposed to detect DICT in the early stage of drug development, but the success is limited as evident by the attrition rate at all clinical phases that remains high due to DICT. Cardiotoxicity can lead to a broad range of heart problems such as arrhythmia, myocardial dysfunction, and QT/QTc prolongation. These problems may be associated with drug properties and drug classes. Therefore, a large drug list with consistent cardiotoxicity annotation is essential to developing methods, such as alternative methodologies and in silico approaches, to improve DICT detection.

Methods: We constructed a DICT classification and annotation scheme to classify FDA-approved drugs regarding their potential to cause DICT based on the FDA approved drug labeling documents. DICT classification scheme involved the extraction of drug labeling documents, establishment of keywords to identify drugs with potential to cause DICT, and the following text-mining and manual reading to curate and generate DICTrank – a large list of drugs with consistent annotation of cardiotoxicity.

Results: DICTrank consists of 1318 drugs that were grouped into four classes: Most-DICT concern (341), Less-DICT concern (528), No-DICT concern (343), and Ambiguous-DICT concern (106; no sufficient information in the labeling document for cardiotoxicity). DICTrank covered a wide range of therapeutic categories, those of a high prevalence in cardiotoxicity, such as sex hormones, cancer drugs and non-steroidal anti-inflammatory drugs.

Conclusions: The developed DICTrank yielded the current largest DICT drug list, which could contribute to the development of new approach methods (NAMs) for the early identification of DICT risk liability during drug development.

Keywords: Drug-induced cardiotoxicity (DICT), FDA Label, Annotation, Drug List.

Poster Number: 26

Characterization and phylogenetic analysis of ADAMs genes in vertebrates.

Vinay Raj, Department of Biology, University of Arkansas at Pine Bluff, Arkansas, U.S.A

Background: Members of a disintegrin metalloproteinases (ADAMs) family regulate diverse cellular functions, including cell adhesion, migration, cellular signaling, and proteolysis. Dysregulation of these processes through aberrant ADAM expression or sustained ADAM activity is linked to chronic inflammation, inflammation-associated cancer, and tumorigenesis. In humans, 22 functional ADAMs have been identified, but only some members of this family have proteolytic activity. The functional roles of mammalian proteolytically active ADAMs are reflected, in part, by their relative amino acid sequence homology and by their tissue distribution.

Methods: The exon and intron structures of the ADAMs genes obtained from ENSEMBL gene annotation information were analyzed. Motif analysis was performed by identifying the conserved motifs of the ADAMs proteins. The sequences of annotated ADAM proteins with proteolytic activity from humans were BLASTed against the EST databases and genome sequences of other vertebrate species. For phylogenetic analyses, deduced amino acid sequences encoded by *ADAMs* genes from various species were aligned to assess the evolutionary relationship among ADAMs.

Results: The study of the physicochemical properties of the ADAMs proteins elucidated the proteolytic activity of the ADAMs proteins. Sequence alignments and phylogenetic analysis provided an understanding of relationships among ADAMs and the broader evolution.

Conclusions: Overall, the results provide valuable information that clarifies the evolutionary relationships of the ADAMs gene family and contributes to the understanding of the biological function of the ADAMs genes.

Keywords: ADAM genes, Proteolytic activity, Physiochemical, Evolution.

Poster Number: 27

Assessment of a Modified Sandwich Estimator for Generalized Estimating Equations with Application to Opioid Poisoning in MIMIC-IV ICU Patients

Paul Rogers[1*], Julie Stoner[2†]

1. Division of Bioinformatics and Biostatistics, National Center for Toxicological Research, U.S. Food & Drug Administration, Jefferson, AR 72079, USA
2. Department of Biostatistics and Epidemiology, College of Public Health, University of Oklahoma Health Sciences Center, Oklahoma City, OK 73126, USA

* Correspondence: paul.rogers@fda.hhs.gov

† Dedicated to the memory of Julie Stoner, a committed public health warrior.

Deceased 18 June 2020.

Background: Longitudinal regression models for correlated binary outcomes are frequently fit using generalized estimating equations (GEE). The Liang and Zeger sandwich estimator is often used in GEE to produce unbiased standard error estimation for regression coefficients in large sample settings, even when the covariance structure is misspecified. The sandwich estimator performs optimally in balanced designs when the number of participants is large with few repeated measurements. However, the sandwich estimator's asymptotic properties do not hold in small sample and rare-event settings. Under these conditions, the sandwich estimator underestimates the variances and is biased downwards.

Method: Here, the performance of a modified sandwich estimator is compared to the traditional Liang-Zeger estimator and alternative forms proposed by authors Morel, Pan, and Mancl-DeRouen. Each estimator's performance was assessed with 95% coverage probabilities for the regression coefficients using simulated data under various combinations of sample sizes and outcome prevalence values with independence and autoregressive correlation structures.

Results: We demonstrated in simulations with sample sizes of 100 subjects and an autoregressive covariance structure with higher correlation settings that all sandwich estimators produced coverage probabilities below 95%. This was not observed in our earlier simulations with low correlation values. As the sample sizes dropped under these same correlation conditions, the Liang-Zeger continued to perform abysmally while the Rogers-Stoner and Pan estimators adjusted. As the sample sizes decreased under a 0.10 correlation with 10% and 5% outcome prevalences, the coverage probabilities of the Liang-Zeger continued to deteriorate, while the Rogers-Stoner and Pan estimators recovered, almost achieving 95% coverage probabilities at 40 subjects and lower.

Conclusion: In our limited simulation settings, the Rogers-Stoner sandwich estimator outperformed the Liang-Zeger, and other estimators as the prevalence and sample size decreased. This approach provides a method for modeling rare events in finite samples on the effects of medications, drugs, and poisons.

Keywords: Sandwich estimator; Generalized estimating equation; Rare event; Finite sample; Binary outcome.

Poster Number: 28

Toward Automation of Diagnosis for Multiple Myeloma using Natural Language Processing

Michael W. Rutherford[1*], Phillip C. Farmer[1,2], Sharmilan Thanendrarajan[2], Jiang Bian[3], Fred W. Prior[1], Christopher P. Wardell[1,2], Jonathan P. Bona[1]

1. Department of Biomedical Informatics, University of Arkansas for Medical Sciences, Little Rock, AR, United States
2. Myeloma Center, University of Arkansas for Medical Sciences, Little Rock, AR, United States
3. Department of Health Outcomes and Bioinformatics, University of Florida, Gainesville, FL, United States

*Corresponding author

Background: With the increased use of advanced medical imaging, updated recommendations from the International Myeloma Working Group (IMWG) for the diagnosis of multiple myeloma now include biomarkers found only in medical imaging. Multiple myeloma, the second most common age-related hematologic malignancy in the US is a disorder that produces bone lesions that are viewable using medical imaging. In this study, we apply Natural Language Processing (NLP) and deep-learning techniques to leverage radiology reports for multiple myeloma diagnosis.

Methods: Our dataset includes more than 15,000 patients and 228,000 imaging studies from the Myeloma Center at the University of Arkansas for Medical Sciences. The set includes multiple modalities, including Computed Tomography (CT), Magnetic Resonance Imaging (MRI), Positron Emission Tomography (PET) and Diffusion Weighted Imaging with Background Suppression (DWIBS).

Our pipeline pre-processes the radiology reports, splitting them into smaller sections including the “findings” and “impressions”. Radiology reports were labeled with diagnoses which occurred just before or after. This resulted in 72,000 labeled reports split into training, testing, and validation sets. Weights and Biases’ Bayes probabilistic search was used to tune hyperparameters for optimal model performance for fine-tuning state of the art transformer models, including BERT, BigBird, and Longformer, to perform sequence classification for predicting multiple myeloma using radiology reports.

Results: Preliminary results show that fine-tuning the models using all radiology reports for classifying any modality achieved the best result with an F1 score of 82.62% for binary classification. PET achieved the best F1 score for modality-specific prediction with 81.86%; followed by MRI with 79.53%, CT with 75.69%, then DWIBS with 71.37%. The categorical classification showed comparable results with slightly lower precision and F1 scores.

Conclusions: Results show that classifying radiology reports for myeloma prediction may be a viable solution for identifying actionable reports requiring priority by clinicians.

Keywords: multiple myeloma, diagnosis, NLP, transformers

Poster Number: 29

Modeling of III-V-on-Sapphire Waveguides for Sapphire-based Photonic Integrated Circuits Platform

Joel Ruzindana, Kennedy Abanihe, Wisdom Arigbofo, Manoj K. Shah*, and Mansour Mortazavi

Department of Chemistry and Physics, University of Arkansas at Pine Bluff, Pine Bluff, AR, USA

*Corresponding author

Background: Photonic integration circuits (PICs) have the potential to deliver a chip with reduced size-weight-power-and-cost. PICs have been demonstrated in various material systems such as III-V, Si, Si_3N_4, $LiNbO_3$ with varying levels of functionality. The thermal expansion mismatch between epitaxial film and substrate is the dominant factor responsible for the generation of large number of defects and eventually, device failure. From a material growth perspective and closely matching linear coefficient of thermal expansion of sapphire to that of GaAs and GaSb, sapphire is a favorable substrate for the growth of III-V materials. Thus, a sapphire-based platform has the potential to be used for large-scale integration platforms just like silicon-based photonic integrated platforms.

Methods: We studied a GaSb/AlSb-on-Sapphire waveguide for Sapphire-based photonic integrated circuit platform by finite-element-method (FEM) using commercial software Ansys. The materials GaSb, AlSb, and Sapphire were used for core, buffer, and substrate layers respectively to design rib and strip waveguides. Using FEM, we numerically investigated multi-mode, single-mode and cut-off conditions and single-mode propagation loss in the GaSb/AlSb-on-sapphire straight waveguides over a broad optical wavelength.

Results: We presented the cut-off, single-mode and multi-mode operation conditions of rib and strip waveguides. The higher index contrast between core and substrate layer allowed us to design compact, low-loss waveguides in the mid-infrared regime.

Conclusion: The presented low-loss, GaSb/AlSb-on-sapphire photonic integrated platform would enable a range of applications in defense systems, and numerous civilian applications such as big data machine learning, fiber optic communication, instrumentation, RF photonics, space exploration, and in nuclear applications.

Keywords: Group III-V materials, optical waveguides, photonic integrated chips, finite-element-method.

Poster Number: 30

Neural Cell Video Synthesis via Optical-Flow Diffusion

Manuel Serna-Aguilera[1], Nathaniel Harris[2], Min Zou[2], Khoa Luu[1]*

1. Computer Science Computer Engineering Department, University of Arkansas, Fayetteville, AR, United States
2. Mechanical Engineering Department, University of Arkansas, Fayetteville, AR, United States

*Corresponding author

Background: Collecting data samples like images and video in the biomedical field can be expensive in terms of time and necessary equipment. This is true for growing and maintaining neuron cells in culture for an extended period of time. This study investigates if the recent video diffusion model can help alleviate this data scarcity problem by synthesizing new videos. This work also investigates how well synthesized subjects are compared to the real data and if learning optical flow can help improve synthesis results.

Methods: The input data are approximately 2,200 small-resolution videos of neuron cells in a culture, imaged via phase contrast microscopy. We compare characteristics of the cells (e.g., body shape, perimeter, and neurite lengths) using two-sample *t*-tests to see if the distributions are significantly similar or not. We approach cell movement and behavior in the synthesized videos qualitatively. To synthesize videos, we use the recent work called video diffusion; it is a deep learning model based on diffusion capable of learning from an input distribution (in our case, the distribution of neuron cell videos).

Results: The baseline method of video diffusion synthesizes cells accurately, although there are some minor but, we argue, significant details that need to be addressed. The general appearance of cells is accurate, however, synthesis details concerning neurites fall short compared to the real data. Integrating optical flow information into the video diffusion model, however, did not improve results as hoped.

Conclusions: The deep learning model of video diffusion is a fantastic tool for synthesizing videos of small subjects such as neuron cells, however, there are several improvements that need to be made to be as true as possible to real data.

Keywords: neurons, video, microscopy, deep learning

Poster Number: 31

Comparative Metagenomic Analysis on Fecal Microbiome of Pregnant Goat

Prajjwal Shrestha[1], Nesreen Aljahdali,[2] Steven Foley[3], Bruce Erikson[3], Monique Felix[1], Yasser M. Sanad[1,3*]

1. Department of Agriculture, University of Arkansas, Pine Bluff, Pine Bluff, AR 71601

2. Biological Science Department, College of Science, King AbdulAziz University, Jeddah, Saudi Arabia

3. FDA National Center for Toxicological Research, Jefferson, AR 72079

*Corresponding author

Background: Food animals, including small ruminants, are a primary source of foodborne infections. Further, manure from food animals can, and do make their way into the local environment where it can lead to contamination of food and water. Consumers are increasingly supporting small ruminant production systems for financial reasons and as a food source, yet limited data is available about the safety of food from these animals. The aim of this study was to examine the impact of pregnancy in goats on the ecology of their gastrointestinal microbiomes and determine the presence and distribution of potential the foodborne bacteria related to pregnancy.

Methods: For that, shotgun and 16S metagenomic sequencing using Illumina MiSeq was performed on the feces collected from 5 does, three that were pregnant and two that were not. The fecal samples were collected once a week over 6 weeks with total 28 samples. The microbiome populations in the pregnant were compared during late pregnancy and after delivery. Sequencing data was analyzed using the MG-RAST analyses pipeline and QIIME-2. **Results:** Overall, the phyla Bacteroidetes and Firmicutes corresponded to 42% and 39% of the taxa detected, respectively. However, after delivery, the relative abundance of Firmicutes increased to approximately 50%. There were approximately ~ 300 variable species identified among all the analyzed samples. Notably, does following delivery had lower abundances of *Campylobacterales* and *Enterococcaceae*.

Conclusions: These data suggest that the pregnancy in small ruminants may play an important factor shaping overall gut microbiomes in goats. This study provides initial data to help better understand the influence of pregnancy in small ruminants and the changes in microbiome structure associated with pregnancy and delivery, which has potential effects to cause foodborne illness.

Keywords: Food animal, Ruminant, Microbiome

Poster Number: 32

New Approach Methods (NAMs) for Drug-induced Liver Injury Prediction: A Comparative Study

Shivangi Shrimali, Weida Tong, Dongying Li*

National Center for Toxicological Research, US Food and Drug Administration, Jefferson, AR, USA 72079

*Corresponding author

Background: One of the most unpredictable adverse effects in human and the main reason for post-marketing drug withdrawals is drug-induced liver injury (DILI). Many studies have been conducted in the previous years utilizing various models to predict DILI potential but often resulted in different conclusions. Specifically, there is no agreement on the most reliable model for predicting DILI risk. There are two factors that need to be considered using New Approach Methods (NAMs), the cost and throughput, a desirable model should be of low cost with high throughput.

Methods: We conducted a comparative analysis of 7 DILI models/assays, ranging from the most inexpensive ones (e.g., in silico approaches) to the most expensive NAMs (e.g., liver-chip). The pairwise comparison was based on the common drugs shared between two models/assays.

Results: We found that the expensive assays didn't exhibit significant advantage over inexpensive ones. Liver chip results were consistent with several in vitro and in silico results but the observation was made over only a small sample size. For a pair of structurally similar drugs (tolcapone vs entacapone) with opposite DILI liability, most of the assays could distinguish the two.

Conclusion: While the sample size is relatively small in this study, it provided insights to future studies to understand the strength and limitations of different models when making decisions for drug development and safety.

Keywords: DILI, drug safety, comparative analysis

Poster Number: 33

Development of the Novel Vitamin E Analogue, Tocoflexol, as a Potent Radiation Countermeasure

Stephen Shrum[1,2], Shivangi Shrimali[1], Awantika Singh[1], Rajeshkumar Manian[2], Nathalie Pineda[1], Saumyadip Nemu[1], David E. Mery[1,2], Shraddha Thakkar[1], Darin E. Jones[1], Nukhet Aykin-Burns[1], Philip Breen[1], and Cesar M. Compadre[1]

1. Department of Pharmaceutical Sciences, University of Arkansas for Medical Sciences, Little Rock, Arkansas USA
2. Tocol Pharmaceuticals LLC, Little Rock, Arkansas USA

* Corresponding author

Background: There is a need to develop safe and effective radiation medical countermeasures that can be used clinically in emergency situations. A major nuclear attack or nuclear reactor accident would result in catastrophic health consequences to millions of people, a danger which is growing by the day. Currently there are no safe and effective radioprotectors and radiomitigators that can offer multi-organ protection, when administered before or after radiation exposure, respectively. Vitamin E tocotrienols have remarkable radioprotective activity when pre-administered subcutaneously, protecting the hematopoietic, gastrointestinal, and other systems against radiation, with very low toxicity. However, tocotrienols have poor pharmacokinetic properties (low oral bioavailability, delayed subcutaneous absorption, and rapid elimination) that render them unsuitable for radiomitigation when administered after radiation.

Methods: To overcome the limitations of the tocotrienols, we have used computational analysis and molecular modeling techniques to develop a synthetic tocotrienol analogue named tocoflexol. Tocoflexol is designed to have an improved pharmacokinetic profile, suitable for radiomitigation while retaining the powerful therapeutic properties of tocotrienols. We characterized the pharmacological properties of tocoflexol using *in vitro* cell models, and then we evaluated its radioprotective efficacy using an irradiated mouse model.

Results: Our computational analysis shows that tocoflexol has superior binding capacity to ATTP, the key transporter that reduces the elimination rate of tocols. Tocoflexol has strong antioxidant properties comparable to tocotrienols, which are key for protecting against radiation, along with rapid cell uptake. Our pharmacokinetic calculations also predict considerably improved oral bioavailability of tocoflexol *in vivo*. Most importantly, we show that subcutaneous administration of tocoflexol (300 mg/kg) in mice 24 hours prior to lethal total-body irradiation (9.5 Gy) is powerfully radioprotective, providing 100% survival.

Conclusions: Tocoflexol is a promising option to develop into an efficacious radioprotectant product to protect populations during radiological emergencies.

Keywords: Radiation, radioprotectors, tocotrienols, computational analysis, molecular modeling

(The University of Arkansas has received patent protection for parts of this research. A potential royalty stream may occur consistent with the University of Arkansas policy.)

Poster Number: 34

Exploring DNA Methylation Changes in Soybean during Infection by *Phytophthora sojae*: Implications for Plant-Pathogen Interactions

Sachleen Singh[1], Shyaron Poudel[2], Dr. Mohamed Milad[3], Dr. Asela Wijeratne[4]*

1. Arkansas Bioscience Institute, Arkansas State University, Jonesboro, AR, US
2. Arkansas Bioscience Institute, Arkansas State University, Jonesboro, AR, US
3. Computer Science & Mathematics, Arkansas State University, Jonesboro, AR, US
4. Arkansas Bioscience Institute and Department of Biological Sciences, Arkansas State University, Jonesboro, AR, US

*Corresponding author: awijeratne@astate.edu

Background: An extremely destructive oomycete pathogen, *Phytophthora sojae*, causes Phytophthora stem and root rot (PSR) in soybeans and can cause extensive economic damages leading to up to $250 million loss annually in the US alone. The main tactic for controlling PSR is using soybean cultivars with genes (*Rps*-genes) that confer resistance to the pathogen. Yet the pathogen is known to rapidly evolve and overcome these genes, making it essential to identify alternative strategies to control the disease. One suggested remedy is understanding the molecular mechanism by which the *Rps* genes control the immune response and altering it to provide more robust resistance. Emerging evidence suggests that DNA methylation is essential in regulating immune responses against invading pathogens. However, its role in soybean-*P. sojae* interaction is unclear.

Methods: Therefore, this study used whole genome sequencing and Enzymatic Methyl Conversion (EM Seq) to identify differentially methylated regions (DMRs) in soybean during *P. sojae* infection.

Results: Our results revealed significant changes in DNA methylation patterns in the infected plants compared to a control.

Conclusions: Overall, our study provides insight into the role of DNA methylation in the soybean-*P. sojae* interaction and sheds light on potential mechanisms underlying plant-pathogen interactions.

Key words: gene regulation, plant-pathogen interactions, immune response

Poster Number: 35

Autoencoder-Based Genotype Imputation for Enhancing Toxicogenomic Data

Meng Song and Huixiao Hong*

Division of Bioinformatics and Biostatistics, National Center for Toxicological Research, US Food and Drug Administration, Jefferson, AR 72079, USA

*Corresponding author

Background: The large-scale human genome sequencing techniques have provided researchers with the opportunities to identify genome polymorphisms associated with heritable traits in response to drugs and environmental chemicals in toxicogenomics. On the other hand, with the mature of multi-omics techniques, toxicogenomics has begun to integrate multiple omics profiles such as transcriptomics, proteomics, and metabolomics to identify the potential molecular biomarkers. However, toxicogenomics faces the problem of missing values, which can considerably obstruct the identification of useful biomarkers of disease and exposure in toxicity analyses. The reasons for the occurrence of missing values include budget limitations, insufficient sample availability, and experimental constraints.

Methods: To address the missing value problem in toxicogenomics, we developed a convolutional autoencoder (AE) model for genotype imputation and implemented a customized training loop by modifying the training process with a single batch loss rather than the average loss over batches used by existing sparse convolutional denoising autoencoder (SCDA) model. This modified AE imputation model was evaluated using a yeast dataset, the human leukocyte antigen (HLA) data from the 1,000 Genomes Project, and the Louisiana Osteoporosis Study (LOS) data.

Results: Our modified AE imputation model achieved comparable or better performance than the SCDA model in terms of evaluation metrics such as the concordance rate (CR), the Hellinger score, the scaled Euclidean norm (SEN) score, and the imputation quality (IQ) score. For the HLA and LOS data, our AE model achieved an average CR of 0.9459 and 0.9005, Hellinger score of 0.9518 and 0.9384, SEN score of 0.9953 and 0.9940, and IQ score of 0.9044 and 0.8681 at missing ratio 20%, respectively.

Conclusions: In summary, our proposed method for genotype imputation has a great potential to facilitate discovery of toxicological mechanisms at the molecular level by increasing data sizes.

Keywords: genotype imputation, deep learning, autoencoder

Disclaimer: This abstract reflects the views of the authors and does not necessarily reflect those of the U.S. Food and Drug Administration.

Poster Number: 36

Improving Early Diagnosis and Treatment Monitoring of Tuberculosis Via Machine Learning Cough Analysis

Chandra Suda

Bentonville High School, Bentonville, Arkansas, USA

Background: Tuberculosis (TB), a bacterial disease mainly affecting the lungs, is the leading infectious cause of mortality worldwide before the COVID-19 pandemic. To prevent TB from spreading within the body, which causes life-threatening complications, timely and effective anti-TB treatment is crucial. Cough, an objective biomarker for TB, is a triage tool that monitors treatment response and regresses with successful therapy. Current gold standards for TB diagnosis are slow or inaccessible, especially in rural areas where TB is most prevalent. In addition, current TB diagnoses with machine learning (ML), like utilizing chest radiographs, are ineffective and do not monitor treatment progression.

Methods: To enable effective diagnosis, I developed a mobile app that analyzes coughs' acoustic epidemiology from smartphones' microphones, using a novel ML architecture, to diagnose TB. The architecture includes a 2D-CNN and XGBoost that was trained on 750K+ cough audio samples worldwide after feature extraction (Mel-spectrograms and MFCCs) and data augmentation (White-noise and IR-convolution). The model used Kaiming initialization, consisting of six convolutional layers with ReLu activation and batch normalization, followed by an AdaptiveAvgPool2d layer. I developed a bi-directional LSTM using periodic cough history, in conjunction with the treatment irregularity algorithm (TIA) which takes in a doctor's recommended treatment drug dosage and timeline, to predictively monitor response to TB treatment.

Results: Using a train-validation-test split of 70%-15%-15%, the 2D-CNN+XGBoost model achieved 88% accuracy and surpassed WHO requirements for specificity and precision as a screening tool. The LSTM and TIA effectively (RMSE<0.28) monitor the body's reaction to anti-TB drugs through changes in cough patterns, allowing the ML model to predict a high risk of treatment or dosage irregularity.

Conclusions: This new early detection of drug irregularity can avert TB relapse, reduce drug-induced liver injury (side-effect), and prevent drug-resistant strains. This project helps to improve TB diagnosis while predictively monitoring treatment.

Keywords: Tuberculosis, Machine-Learning, Cough, Treatment Monitoring

Poster Number: 37

CROVIA: Seeing Drone Scenes from Car Perspective via Cross-View Adaptation

Thanh-Dat Truong[1*], Chi Nhan Duong[2], Ashley Dowling[3], Son Lam Phung[4], Jackson Cothren[5], Khoa Luu[1]

1. Department of Computer Science and Computer Engineering, University of Arkansas, Fayetteville, Arkansas, USA
2. Department of Computer Science and Software Engineering, Concordia University, Montreal, Canada
3. Department of Entomology & Plant Pathology, University of Arkansas, Fayetteville, Arkansas, USA
4. Department of Engineering and Information Sciences, University of Wollongong, Northfields Avenue, Wollongong, Australia
5. Department of Geosciences, University of Arkansas, Fayetteville, Arkansas, USA

*Corresponding author

Background: Understanding semantic scene segmentation of urban scenes captured from the Unmanned Aerial Vehicles (UAV) perspective is an important factor in building a perception model for UAV. With the limited large-scale densely labeled data, develop a supervised model trained on UAV data is not an optimal solution and semantic scene segmentation for UAV views also requires a broad understanding of an object from both its top and side views. Adapting knowledge learned from well-annotated autonomous driving data to unlabeled UAV data could be considered as a cost-effective solution but is challenging due to the cross-view differences between the two data types.

Methods: Therefore, our paper introduces a novel Cross-View Adaptation (CROVIA) approach to effectively adapt the knowledge learned from on-road vehicle views to UAV views. In particular, the novel geometry-based constraint to cross-view adaptation is proposed based on the geometric correlation across views. Then, the geometry-based constraint further is derived into the new Geometry-Constraint Cross-View (GeiCo) loss to effectively transferred the cross-view correlation in the image space to the segmentation space. Moreover, the multi-modal bijective networks are proposed to enforce the cross-view global structural modeling.

Results: The results of our experiments have analyzed the effectiveness of our approach in cross-view learning and shown the performance improvement. Particularly, our experimental results on new cross-view adaptation benchmark introduced in this work, i.e., GTA5 → UAVID, show our State-of-the-Art (SOTA) results and outperform prior adaptation methods.

Conclusions: To the best of our knowledge, our work is one of the first studies addressing the cross-view adaptation learning in semantic scene segmentation. Our solution provides a comprehensive, cost-effective approach to transferring knowledge between two different views, thus, improving the performance of the segmentation models.

Keywords: Semantic Segmentation, Cross-View Adaptation, Geometric Constraint.

Poster Number: 38

Evaluating Drug Promiscuity and Preclinical Safety Using *In Vitro* Cytotoxicity Data in HepG2 and PBMC Cell Models

Andy Vo, Terry van Vleet*

Development Biological Sciences, AbbVie, North Chicago, IL, United States.

*Corresponding Author

Background: In the drug discovery and development pipeline, it is becoming increasingly important to identify early indicators of potential safety outcomes. *In vitro* cytotoxicity assessment provides a multiparametric measurement to infer some potential mechanisms of toxicity from compound treatment at the cell level. These measurements are large scale and high throughput in nature, which has resulted in a sizeable database to generalize findings across compounds. Despite this, there has been a large limitation to fully leverage these large datasets due to our lack of understanding of translatability between specific *in vitro* and *in vivo* outcomes. These limitations have been largely due to lack of data harmonization and inoperability between cross-functional datasets.

Methods: In this study, we look to utilize and harmonize in vitro cytotoxicity data in HepG2 and PBMC cells related to approximately 45,000 compounds and understand their relationship to: (1) Chemical structural properties, (2) Bioactivity interaction panels), and (3) *In Vivo* preclinical safety data. In order to achieve this, we applied a variety of machine-learning and rule-based methods to standardize and impute cytotoxicity measurements with HepG2 and PBMC cytotoxicity data, respectively.

Results: These datasets were further harmonized and connected to relevant safety datasets using both domain knowledge and computational enabled workflows to identify significant pairwise relationships across >200 safety features and have identified significant relationships between cytotoxicity in HepG2 and PBMC cells to various clinical pathology and chemistry endpoints.

Conclusions: In summary, we hope to further utilize these relationships to identify relational scores to better predict toxicology endpoints early in the drug development pipeline.

Keywords: Cytotoxicity, Preclinical Safety, Drug Promiscuity

Poster Number: 39

Using Language Model to Facilitate COVID-19-Associated Neurological Disorder Literature Analysis: a BERTox research

Leihong Wu*, Joshua Xu, Weida Tong

Division of Bioinformatics and Biostatistics, National Center for Toxicological Research, 3900 NCTR Rd, Jefferson AR, USA

*Corresponding Author

Background: COVID-19 can lead to multiple severe outcomes including neurological and psychological impacts. However, it is challenging to manually scan hundreds of thousands of COVID-19 articles on a regular basis. To update our knowledge, provide sound science to the public, and communicate effectively, it is critical to have an efficient means of following the most current published data.

Methods: We have launched an AI program at FDA to apply the modern Ai technology in toxicology, called AI4Tox. One of the AI4TOX initiatives is develop the most advanced AI-powered Natural Language Processing (NLP) to facilitate analysis of FDA documents and public literature for improved efficiency and accuracy of information retrieval and toxicity assessment. In this study, we developed a language model to search abstracts using the most advanced artificial intelligence (AI) to accurately retrieve articles on COVID-19-associated neurological disorders. We applied this NeuroCORD model to the largest benchmark dataset of COVID-19, CORD-19.

Results: We found that the model developed on the training set yielded 94% prediction accuracy on the test set. This result was subsequently verified by two experts in the field. In addition, when applied to 96,000 non-labeled articles that were published after 2020, the NeuroCORD model accurately identified approximately 3% of them to be relevant for the study of COVID-19-associated neurological disorders, while only 0.5% were retrieved using conventional keyword searching.

Conclusions: NeuroCORD provides an opportunity to profile neurological disorders resulting from COVID-19 in a rapid and efficient fashion, and its general framework could be used to study other COVID-19-related emerging health issues.

Poster Number: 40

Development of Random Forest Model for Predicting SARS-CoV-2 Main Protease Binders as Potential Candidates for Repurposing to COVID-19 Treatment

Liang Xu, Jie Liu, Minjun Chen, and Huixiao Hong*

Division of Bioinformatics and Biostatistics, National Center for Toxicological Research, U.S. Food and Drug Administration, 3900 NCTR Road, Jefferson, AR 72079, USA.

*Corresponding author

Background: COVID-19 is a global pandemic with millions of people infected. Although US FDA approved several drugs for the treatment of COVID-19, effective COVID-19 treatment drugs are in urgent need. The main protease of SARS-CoV-2 is a major target for COVID-19 drugs, and repurposing FDA approved drugs that bind the main protease of SARS-CoV-2 could be potential candidates for the treatment of COVID-19.

Methods: We collected 372 ligands from the 3D structures of the ligand-bound main protease of SARS-CoV-2 from the protein data bank and used as binders for training. We then curated 259 compounds experimentally tested in SARS-CoV-2 main protease binding assays from the literature. Of the curated non-binders, 187 were used for training. The rest compounds (both binders and non-binders) were used for testing. We curated 1284 FDA-approved drugs from diverse sources including drug labeling documents for identification of SARS-CoV-2 main protease binders for repurposing. Random forest algorithm was used for constructing predictive models based on molecular descriptors calculated using Mold2 software. Model performance was evaluated using 500 iterations of 5-fold cross validations and the testing data set.

Results: The random forest models showed 84.2% and 78.6% prediction accuracy in the 5-fold cross validations and the testing, respectively. The random forest model constructed from the whole training data set was used to predict SARS-CoV-2 main protease binders as potential candidates for repurposing to COVID-19 treatment.

Conclusions: Our results demonstrate that machine learning could be an efficient method for drug repurposing, and thus accelerate the drug development targeting SARS-CoV-2.

Keywords: COVID-19, SARS-Cov-2, main protease, machine learning, random forests, drug

Disclaimer: This abstract reflects the views of the authors and does not necessarily reflect those of the U.S. Food and Drug Administration.