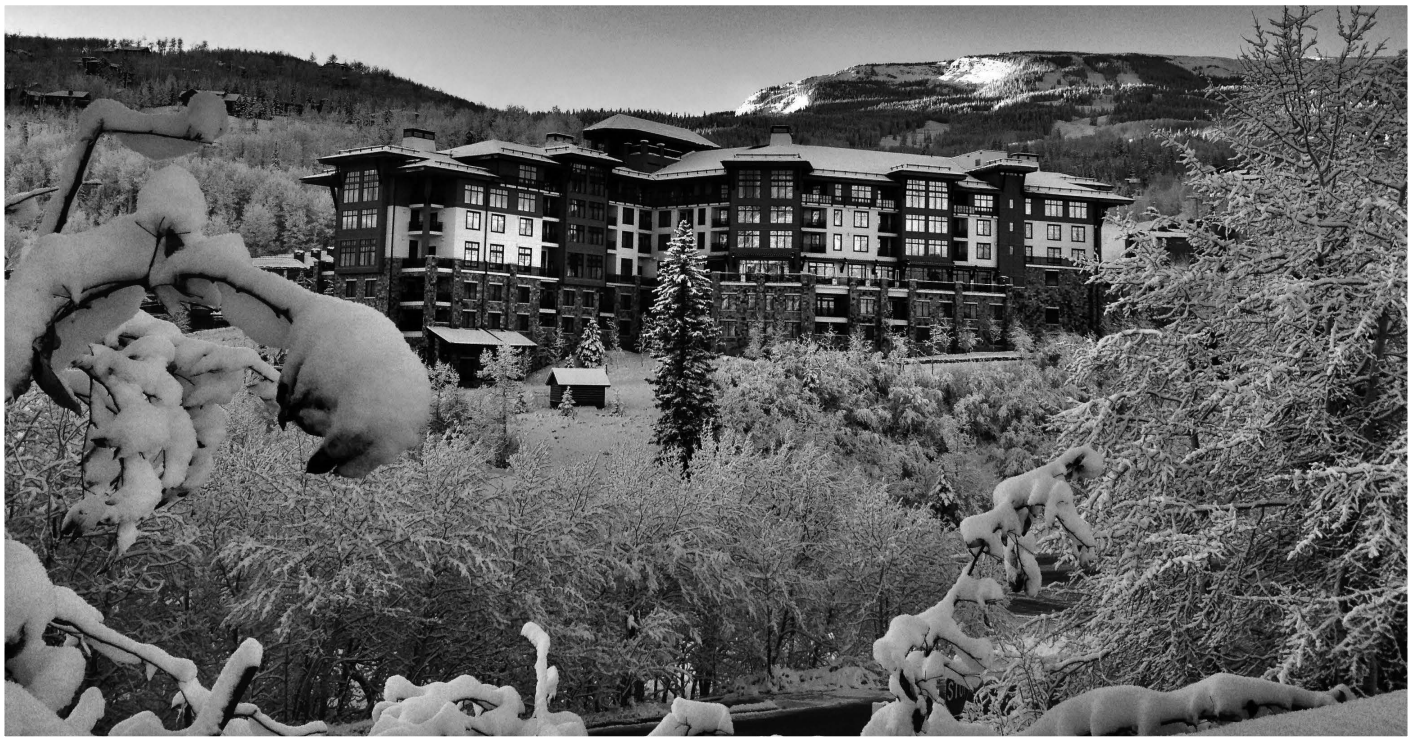
SEVENTEENTH ROCKY MOUNTAIN BIOINFORMATICS CONFERENCE

ROCKY 2019
Bioinformatics Conference



ISCB INTERNATIONAL SOCIETY FOR COMPUTATIONAL BIOLOGY

DECEMBER 5 TO 7, 2019

SNOWMASS/ASPEN
COLORADO

Conference Chair
Lawrence Hunter, PhD
University of Colorado Anschutz
Medical Campus,
Computational Bioscience Program

Conference Sponsors

GOLD SPONSORS

IBM
PRECISION quartzbio℠
SomaLogic

SILVER SPONSOR

biodesix

AGENDA AT-A-GLANCE

WEDNESDAY – December 4, 2019

TIME	SESSION TYPE
4:00 PM – 6:00 PM	REGISTRATION

THURSDAY – December 5, 2019

TIME	SESSION TYPE
8:00 AM–6:00 PM	REGISTRATION
8:00 AM–9:00 AM	BREAKFAST
9:00 AM–9:45 AM	**KEYNOTE 1: KEVIN BRETONNEL COHEN**, *PhD, Director, Biomedical Text Mining Group Computational Bioscience Program, University of Colorado School of Medicine, USA* **Two Existential Threats to Biomedical Text Mining...and How to Address Them with Natural Language Processing**
9:45 AM–10:25 AM	OP 01 - OP 04
10:25 AM–10:45 AM	BREAK
10:45 AM–11:15 AM	**KEYNOTE 2: JOSEPH ALLISON**, *PhD, Bioinformatic Scientist, SomaLogic, Inc., Colorado, USA* **A Story of Proteomic Statistical Process Control at SomaLogic**
11:15 AM–11:55 AM	OP 05 - OP 08
12:00 PM–4:00 PM	SKI BREAK
4:00 PM–4:30 PM	**KEYNOTE 3: OLUWATOSIN OLUWADARE**, *PhD, Assistant Professor, Department of Computer Science and Bachelor of Innovation, College of Engineering and Applied Science, University of Colorado, USA* **3D Chromosome and Genome Structure Modeling**
4:30 PM–5:10 PM	OP 09 - OP 12
5:10 PM–5:30 PM	BREAK
5:30 PM–6:00 PM	OP 13 - OP 15
6:00 PM – 6:30 PM	Open ballroom
6:30 PM–9:30 PM	DINNER — IL POGGIO

FRIDAY – December 6, 2019

TIME	SESSION TYPE
8:00 AM – 6:00 PM	REGISTRATION
8:00 AM – 9:00 AM	BREAKFAST
9:00 AM – 9:45 AM	**KEYNOTE 4: JOYCE C. HO**, *PhD Assistant Professor, Computer Science Department, Emory University, Georgia, USA* **Developing an Evidence Matching Framework Using Web-based Medical Literature**
9:45 AM – 10:25 AM	OP 16 - OP 19
10:25 AM – 10:45 AM	BREAK
10:45 AM – 11:15 AM	**KEYNOTE 5: NIMISHA SCHNEIDER,** *PhD* and **TED FOSS**, *PhD QuartzBio, part of Precision for Medicine Maryland, USA* **Coupling Data-Driven and Mechanistic Modeling Approaches Through the Application of a Scalable, Knowledge-Driven Framework and High-Throughput Public Omics Data Sources**
11:15 AM – 11:55 AM	OP 20 - OP 23
12:00 PM – 4:00 PM	SKI BREAK
4:00 PM – 4:30 PM	**KEYNOTE 6: JUSTIN GUINNEY**, *PhD, VP, Computational Oncology, Sage Bionetworks, Affiliate Associate Professor, University of Washington, Director, DREAM Challenges, Washington, USA* **The Model-to-Data Paradigm: Overcoming Data Access Barriers in Biomedical Competitions**
4:30 PM – 5:10 PM	OP 24 - OP 27
5:10 PM – 5:30 PM	BREAK
5:30 PM – 6:00 PM	OP 28 - OP 30
6:00 PM – 6:30 PM	Open ballroom
6:30 PM – 8:30 PM	POSTER SESSION

SATURDAY – December 7, 2019

TIME	SESSION TYPE
8:00 AM – 11:00 AM	REGISTRATION
8:00 AM – 9:00 AM	BREAKFAST
9:00 AM – 9:30 AM	**KEYNOTE 7: HEINRICH RÖDER**, *PhD, Biodesix, Inc., Colorado, USA* **Development of Clinically Relevant Tests from Human Serum Samples: A Look at the Circulating Proteome**
9:30 AM – 10:20 AM	OP31 – OP35
10:20 AM – 10:40 AM	BREAK
10:40 AM – 11:30 AM	OP36 – OP40
11:30 AM – 12:00 PM	**KEYNOTE 8: KIRK E. JORDAN**, *PhD, IBM Distinguished Engineer, Data Centric Solutions, IBM T.J. Watson, Research, and Chief Science Officer, IBM Research UK* **Intelligent Simulations — Incorporating AI into Computational Simulations**
12:00 PM – 12:15 PM	CLOSING RAFFLE AND AWARDS

DETAILED AGENDA

WEDNESDAY – December 4, 2019

TIME	SESSION TYPE
4:00 PM – 6:00 PM	REGISTRATION

THURSDAY – December 5, 2019

TIME	SESSION TYPE
8:00 AM – 6:00 PM	**REGISTRATION**
8:00 AM – 9:00 AM	**BREAKFAST**
9:00 AM – 9:45 AM	**KEYNOTE 1: KEVIN BRETONNEL COHEN**, *PhD* *Director, Biomedical Text Mining Group, Computational Bioscience Program, University of Colorado School of Medicine, USA* **Two Existential Threats to Biomedical Text Mining...and How to Address Them with Natural Language Processing**
9:45 AM – 9:55 AM	**OP 01: Next Generation Sequencing as a preventive medicine tool: A Mexico-United States collaboration.** *Presenting Author: Alfredo Gilberto de la Garza-Galvan, University of Texas Health Science Center at Houston* CANCELLED
9:55 AM – 10:05 AM	**OP 02: Enabling the next generation of microbiome science with QIIME 2** *Presenting Author: Greg Caporaso, Northern Arizona University*
10:05 AM – 10:15 AM	**OP 03: Evolutionary Action is a unifying framework for assessing missense variant structures within and across phyla** *Presenting Author: Nicholas Abel, Baylor College of Medicine*
10:15 AM – 10:25 AM	**OP 04: A toxicogenomics approach to identify liver and kidney injuries** *Presenting Author: Patric Schyman, Biotechnology HPC Software Applications Institute (BHSAI)*
10:25 AM – 10:45 AM	**BREAK**
10:45 AM – 11:15 AM	**KEYNOTE 2: JOSEPH ALLISON**, *PhD, Bioinformatic Scientist , SomaLogic, Inc. , Colorado, USA* **A Story of Proteomic Statistical Process Control at SomaLogic**
11:15 AM – 11:25 AM	**OP 05: Towards Automating Computational Phenotyping: Exploring the Trade-offs of Different Vocabulary Mapping Strategies** *Presenting Author: Tiffany Callahan, University of Colorado Denver Anschutz Medical Campus- Computational Bioscience Program*
11:25 AM – 11:35 PM	**OP 06: Open PBTA: Collaborative analysis of the Pediatric Brain Tumor Atlas** *Presenting Author: Joshua Shapiro, Childhood Cancer Data Lab (Alex's Lemonade Stand Foundation)*
11:35 PM – 11:45 PM	**OP 07: Data Discovery Engine: A web-based toolset for maximizing data discoverability and promoting reusable data-sharing best practices** *Presenting Author: Marco Cano, Scripps Research*
11:45 PM – 11:55 PM	**OP 08: Hypergraph Analytics for Computational Virology** *Presenting Author: Cliff Joslyn, Pacific Northwest National Laboratory*
12:00 PM – 4:00 PM	**SKI BREAK**
4:00 PM – 4:30 PM	**KEYNOTE 3: OLUWATOSIN OLUWADARE**, *PhD* *Assistant Professor, Department of Computer Science and Bachelor of Innovation, College of Engineering and Applied Science, University of Colorado, USA* **3D Chromosome and Genome Structure Modeling**
4:30 PM – 4:40 PM	**OP 09: Containerized pipeline for the identification of compound heterozygous variants in trios** *Presenting Author: Dustin Miller, Brigham Young University*
4:40 PM – 4:50 PM	**OP 10: The design of an interactive lung map for studying premalignant lesions in the lung over time** *Presenting Author: Carsten Görg, Colorado School of Public Health*
4:50 PM – 5:00 PM	**OP 11: Apollo: an efficient tool to collaboratively refine and attribute genome-wide genomic annotationsc Annotations** *Presenting Author: Nathan Dunn, University of California, Berkeley*
5:00 PM – 5:10 PM	**OP 12: An online end-to-end pipeline for virus phylogeography that leverages Natural Language Processing for finding host locations** *Presenting Author: Matthew Scotch, Arizona State University*
5:10 PM – 5:30 PM	**BREAK**
5:30 PM – 5:40 PM	**OP 13: Evolutionary Action as a Tool for Quantifying Differentiation Across the Primate Family Tree** *Presenting Author: Harley Peters, Baylor College of Medicine*
5:40 PM – 5:50 PM	**OP 14: Haplocravat: LD-based calculation built on top of a platform to annotate variants** *Presenting Author: Ben Busby, Mountain Genomics/Johns Hopkins University*
5:50 PM – 6:00 PM	**OP 15: Characterizing the Regulatory Framework in an Aggressive Breast Cancer Phenotype: A Bayesian Regression-Based Enhancement** *Presenting Author: George Acquaah-Mensah, Massachusetts College of Pharmacy & Health Sciences*
6:30 PM – 9:30 PM	**DINNER - IL POGGIO**

DETAILED AGENDA

FRIDAY – December 6, 2019

TIME	SESSION TYPE
8:00 AM – 6:00 PM	**REGISTRATION**
8:00 AM – 9:00 AM	**BREAKFAST**
9:00 AM – 9:45 AM	**KEYNOTE 4: JOYCE C. HO**, *PhD , Assistant Professor , Computer Science Department , Emory University , Georgia, USA* **Developing an Evidence Matching Framework Using Web-based Medical Literature**
9:45 AM – 9:55 AM	**OP 16: Discovering Subclones in Tumors Sequenced at Standard Depths** *Presenting Author: Li Liu, Arizona State University*
9:55 AM – 10:05 AM	**OP 17: Pathogenic Synonymous Variants Are More Likely to Affect Codon Usage Biases than Benign Synonymous Variants** *Presenting Author: Justin Miller, Brigham Young University*
10:05 AM – 10:15 AM	**OP 18: Map and model — moving from observation to prediction in toxicogenomics** *Presenting Author: Wibke Busch, Helmholtz Centre for Environmental Research - UFZ*
10:15 AM – 10:25 AM	**OP 19: Combining the Evolutionary Trace Algorithm and Covariation Metrics Yields Improved Structural Predictions** *Presenting Author: Daniel Konecki, Baylor College of Medicine*
10:25 AM – 10:45 AM	**BREAK**
10:45 AM – 11:15 AM	**KEYNOTE 5; NIMISHA SCHNEIDER**, *PhD* *QuartzBio, part of Precision for Medicine, Maryland, USA* **TED FOSS**, *PhD* *Director, Systems and Data Integration Precision For Medicine* *QuartzBio, part of Precision for Medicine Maryland, USA* **Coupling Data-Driven and Mechanistic Modeling Approaches Through the Application of a Scalable, Knowledge-Driven Framework and High-Throughput Public Omics Data Sources**
11:15 AM – 11:25 AM	**OP 20: Toxicological Mechanistic Inference: Generating mechanistic explanations of adverse outcomes** *Presenting Author: Ignacio Tripodi, University of Colorado, Boulder*
11:25 AM – 11:35 PM	**OP 21: Giving credit where credit is due: How to make more meaningful connections between people and their roles, work, and impact** *Presenting Author: Nicole Vasilevsky, Oregon Health & Science University*
11:35 PM – 11:45 PM	**OP 22: Nearest-neighbor Projected-Distance Regression to detect network interactions and control for confounders and multiple testing** *Presenting Author: Trang Le, University of Pennsylvania*
11:45 PM – 11:55 PM	**OP 23: Deep learning enables in silico chemical-effect prediction** *Presenting Author: Jana Schor, Helmholtz Centre for Environmental Research - UFZ*
12:00 PM – 4:00 PM	**SKI BREAK**
4:00 PM – 4:30 PM	**KEYNOTE 6: JUSTIN GUINNEY**, *PhD* *VP, Computational Oncology, Sage Bionetworks, Affiliate Associate Professor, University of Washington, Director, DREAM Challenges, Washington, USA* **The Model-to-Data Paradigm: Overcoming Data Access Barriers in Biomedical Competitions**
4:30 PM – 4:40 PM	**OP 24: Correlations and curses of dimensionality: optimizing k in nearest-neighbor feature selection** *Presenting Author: Bryan Dawkins, University of Tulsa*
4:40 PM – 4:50 PM	**OP 25: Deep Learning based Multi-view model for deciphering gene regulatory keywords** *Presenting Author: Pramod Bharadwaj Chandrashekar, Arizona State University*
4:50 PM – 5:00 PM	**OP 26: Identifying optimal mouse models for human asthma using a novel modeling approach** *Presenting Author: Yihsuan Tsai, University of North Carolina at Chapel Hill*
5:00 PM – 5:10 PM	**OP 27: Bridging the Bioinformatics Knowledge Gap in the Pediatric Cancer Research Community with the Childhood Cancer Data Lab workshops** *Presenting Author: Chante Bethell, Childhood Cancer Data Lab (Alex's Lemonade Stand Foundation)*
5:10 PM – 5:30 PM	**BREAK**
5:30 PM – 5:40 PM	**OP 28: Hetnet connectivity search provides rapid insights into how two biomedical entities are related** *Presenting Author: Daniel Himmelstein, University of Pennsylvania*
5:40 PM – 5:50 PM	**OP 29: Landmark and Cancer-Relevant Gene Selection of RNA Sequencing Data for Survival Analysis** *Presenting Author: Carly Clayman, Penn State University - Great Valley*
5:50 PM – 6:00 PM	**OP 30: The impact of undesired technical variability on large-scale data compendia** *Presenting Author: Alexandra Lee, Genomics and Computational Biology Graduate Program, University of Pennsylvania, Philadelphia, PA, USA; Department of Systems Pharmacology and Translational Therapeutics, University of Pennsylvania,*
6:30 PM – 8:30 PM	**POSTER SESSION**

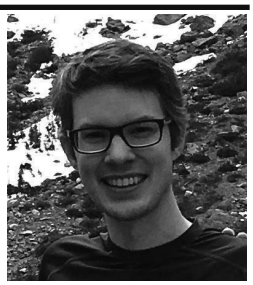
DETAILED AGENDA

SATURDAY – December 7, 2019

TIME	SESSION TYPE
8:00 AM – 11:00 AM	**REGISTRATION**
8:00 AM – 9:00 AM	**BREAKFAST**
9:00 AM – 9:30 AM	**KEYNOTE 7: HEINRICH RÖDER**, *PhD , Biodesix, Inc. , Colorado, USA* **Development of Clinically Relevant Tests from Human Serum Samples: A Look at the Circulating Proteome**
9:30 AM – 9:40 AM	**OP 31: A pedigree-level examination of Schistosoma japonicum following schistosomiasis reemergence in rural China** *Presenting Author: Laura Timm, University of Colorado - Anschutz Medical Campus*
9:40 AM – 9:50 AM	**OP 32: High Resolution Proteomics and Genomics of CNDP1 Repeat Variants Linked to Diabetic Nephropathy** *Presenting Author: Nicholas Kinney, Edward Via College of Osteopathic Medicine*
9:50 AM – 10:00 AM	**OP 33: Computational Analysis of Kinesin Mutations Implicated in Hereditary Spastic Paraplegias** *Presenting Author: Shaolei Teng, Howard University*
10:00 AM – 10:10 AM	**OP 34: Enabling structure-based data-driven selection of targets for cancer immunotherapy** *Presenting Author: Dinler Antunes, Rice University*
10:10 AM – 10:20 AM	**OP 35: Biotherapeutic Protein Immunogenicity Risk Assessment with TCPro** *Presenting Author: Osman Yogurtcu, FDA*
10:20 AM – 10:40 AM	**BREAK**
10:40 AM – 10:50 AM	**OP 36: Large-scale phylogenetic analysis reveals different sequence divergence patterns in orthologous and paralogous proteins** *Presenting Author: Joseph Ahrens, University of Colorado Denver Anschutz Medical Campus*
10:50 AM – 11:00 AM	**OP 37: A new framework for clustering single cell RNA-seq data** *Presenting Author: Ziyou Ren, Northwestern University*
11:00 AM – 11:10 AM	**OP 38: How to annotate what we don't know** *Presenting Author: Mayla Boguslav, University of Colorado Anschutz Medical Campus*
11:10 AM – 11:20 AM	**OP 39: Filtering, classification, and selection of new knowledge for model assembly and extension** *Presenting Author: Natasa Miskov-Zivanov, University of Pittsburgh*
11:20 AM – 11:30 AM	**OP 40: Pathway-based Single-Cell RNA-Seq Classification, Clustering, and Construction of Gene-Gene Interactions Networks Using Random Forests** *Presenting Author: Herbert Pang, University of Hong Kong*
11:30 AM – 12:00 PM	**KEYNOTE 8: KIRK E. JORDAN**, *PhD IBM Distinguished Engineer, Data Centric Solutions, IBM T.J. Watson Research and Chief Science Officer, IBM Research UK* **Intelligent Simulations - Incorporating AI into Computational Simulations**
12:00 PM – 12:15 PM	**CLOSING RAFFLE AND AWARDS**

KEYNOTE SPEAKERS

JOSEPH ALLISON, PhD



Bioinformatic Scientist
SomaLogic, Inc.
Colorado, USA

A Story of Proteomic Statistical Process Control at SomaLogic

The processes that have been implemented to support and run the SomaScan® Assay at SomaLogic are incredibly robust but, before the introduction of Umbrella, post-hoc tracking of these processes and their longitudinal stability was spread across the collective consciousness. Umbrella is a bespoke full-stack data science platform designed to coalesce that collective knowledge and report back to the company at large. These reports range from simple control charting to unsupervised ML pipelines to detect nuanced assay artifact, identify potential causes, and document the impact on our SOMAmer® reagents' signals. It is this system that enables us to be able to robustly distinguish between biological signal and assay artifact.

KEVIN BRETONNEL COHEN, PhD

Director, Biomedical Text Mining Group
Computational Bioscience Program
University of Colorado School of Medicine
USA

Two Existential Threats to Biomedical Text Mining... and How to Address Them with Natural Language Processing

The reproducibility crisis calls into question some of the most fundamental use cases of biomedical natural language processing: if 65% of the scientific literature is questionable, what is the point of mining it? Meanwhile, computational research has mostly been immune to the crisis, but there is no a priori reason to expect that state of affairs to continue. This talk proposes that natural language processing itself can address this issue on both fronts--but how?

JUSTIN GUINNEY, PhD

VP, Computational Oncology, Sage Bionetworks
Affiliate Associate Professor, University of Washington
Director, DREAM Challenges
Washington, USA

The Model-to-Data Paradigm: Overcoming Data Access Barriers in Biomedical Competitions

Data competitions often rely on the physical distribution of data to challenge participants, a significant limitation given that much data is proprietary, sensitive, and often non-shareable. To address this, the DREAM Challenges has advanced a challenge framework called model-to-data (M2D), requiring participants to submit re-runnable algorithms instead of model predictions. The DREAM organization has successfully completed multiple M2D-based challenges, and is expanding this approach to unlock highly sensitive and non-distributable human data for use in biomedical data challenges.

The EHR DREAM Challenges is the first M2D data utilizing EHR data. We are asking participants to predict patient mortality within six months of their last hospital visit, using data from the University of Washington Medical System enterprise data warehouse (2009-2019) with over 1.3 million patients and 22 million visits. Given the highly sensitive nature of EHR data and risks of re-identifiability, the M2D approach is being used to ensure data privacy and protections.

Prior to launching the EHR Challenge, we completed a feasibility study and developed 3 models: demographic information; demographic information and 4 common chronic diseases, and demographic information and the top 20 indications. Model performance using area under the receiver-operating-curve was 0.682, 0.794, and 0.723, respectively. This demonstrated technical robustness of the challenge architecture, and the ability to generate and evaluate predictive algorithms in a secure manner. The EHR Challenge was launched in September 2019, and will close in January 2020.

KEYNOTE SPEAKERS

JOYCE C. HO, PhD

Assistant Professor, Computer Science Department
Emory University
Georgia, USA

Developing an Evidence Matching Framework Using Web-based Medical Literature

Researchers are discovering new disease subgroups from secondary analyses of electronic health records. However, such subgroups need to be validated or aligned with current literature. We developed a scalable framework that produces evidence sets (or relevant articles) using a large corpus of online medical literature. In this talk, I will discuss some of the challenges associated with term representation and mining biomedical text. I will also present a case study of our framework to validate EHR-based phenotypes.

KIRK E. JORDAN, Ph.D.

IBM Distinguished Engineer
Data Centric Solutions
IBM T.J. Watson Research
and
Chief Science Officer
IBM Research UK

Intelligent Simulations - Incorporating AI into Computational Simulations

We are at an inflexion point in the way we are computing today and the way we will compute tomorrow. While computer systems will consist of heterogeneous components adding to complexity we need to make high end systems much more accessible. Combine this machine complexity with the fact that more compute is required to tackle the ever increasing complexity of the problems and data that are driving science investigations especially in the life sciences. How will we address this ever increasing complexity. IBM Research is looking at the Future of Computing through a new lens which consists of bits, neurons and qubits. In this talk I will describe how IBM Research is looking at the future of computing and I will outline how we are increasing the use of Artificial Intelligence (AI) and Machine Learning (ML) to address some of the complexity while enabling us to tackle the interesting science questions we and our collaborators are currently investigating with some examples in the Life Sciences.

OLUWATOSIN OLUWADARE, PhD

Assistant Professor
Department of Computer Science and Bachelor of Innovation
College of Engineering and Applied Science
University of Colorado
USA

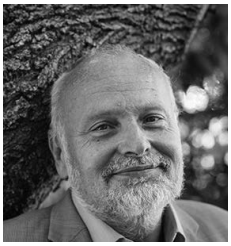
3D Chromosome and Genome Structure Modeling

To improve the understanding of chromosome organization within a cell, chromosome conformation capture techniques such as 3C, 4C, 5C, and Hi-C have been developed. These technologies help to determine the spatial positioning and interaction of genes and chromosome regions within a genome. Using next-generation sequencing strategies such as high-throughput and parallel sequencing, Hi-C can profile read pair interactions on an "all-versus-all" basis—that is, it can profile interactions for all read pairs in an entire genome. The development of chromosome conformation capture techniques, particularly Hi-C, has substantially benefited the study of the spatial proximity, interaction, and genome organization of several cells. In recent years, numerous genome structure construction algorithms have been developed to explain the roles of three-dimensional (3D) structure reconstruction in the cell and to explain abnormalities occurring in a diseased and a normal cell. Three-dimensional inference involves the reconstruction of a genome's 3D structure or (in some cases) ensemble of structures from contact interaction frequencies represented in a two-dimensional matrix. To solve this 3D inference problem, we developed an optimization-based algorithm that performed better than any existing tool for chromosome and genome 3D structure prediction called 3DMax. 3DMax has been packaged as a software tool and is publicly available to the research community. 3DMax performs well with noisy data; also, its performance is unaffected by changing normalization methods, which is not the case for many other existing methods.

KEYNOTE SPEAKERS

HEINRICH RÖDER, PhD

Biodesix, Inc.
Colorado, USA



Development of Clinically Relevant Tests from Human Serum Samples: A Look at the Circulating Proteome

Therapeutics for the treatment of cancer patients have been transformed with the introduction of immunotherapies, starting with checkpoint inhibition. Instead of targeting the tumor itself, the mechanism of action for immunotherapies relies on reactivation of the immune system such that the host can re-engage the cancer using the complex system evolutionarily developed to heal human disease. In oncology, clinical trials have proven that immunotherapies are effective at reducing tumor burden and extending survival in cancer patients across many indications. However, not all patients benefit from all immunotherapies. Specifically, there is a subgroup of patients whose lack of response may be attributed to a compromised immune system referred to as primary immunotherapy resistance (PIR). A test identifying patients with PIR, prior to treatment with specific immunotherapies, would be useful for guiding therapeutic decision making.

Biodesix uses a hypothesis-free approach to build clinically relevant tests allowing the creation of multivariate classifiers related to deep learning that reflect the complexity of biological interactions without any bias from expectations about their mechanisms. Mass spectral data collected from human serum samples are analyzed by the Diagnostic Cortex® robust data analytics machine learning based platform to design classifiers with clinical relevance. Using this approach, we have developed multiple independently validated tests to identify patients with melanoma and lung cancer that have particular poor outcomes on anti-PD1 immunotherapy and therefore may be unsuitable candidates for treatment with checkpoint inhibition.

These tests stratify patients into different immunological phenotypes with different outcomes on immunotherapies. We applied ideas similar to GSEA (Gene Set Enrichment Analysis) to mass spectral data (PSEA: Protein Set Enrichment Analysis) to gain biological insight into the processes detectable in the circulating proteome related to these phenotypes. We found host immunological functions, such as acute phase response, wound healing, and complement system are related to test classification labels indicating that treatment success does not depend solely on a single molecule, protein, or signaling pathway. Our systems biology approach combining proteomics and machine learning methods is hypothesis generating and requires further external validation; however, our findings have been supported by independent research groups using orthogonal approaches.

NIMISHA SCHNEIDER, PhD

QuartzBio, part of Precision for Medicine
Maryland, USA

TED FOSS, PhD

Director, Systems and Data Integration
Precision For Medicine
QuartzBio, part of Precision for Medicine
Maryland, USA

Coupling Data-Driven and Mechanistic Modeling Approaches Through the Application of a Scalable, Knowledge-Driven Framework and High-Throughput Public Omics Data Sources

Advances in high throughput measurement technologies (-omics data) have made it possible, and increasingly affordable, to generate high complexity, high volume data for medical research, and these data are increasingly available to researchers through public sources. Rapidly mining these data for useful mechanistic insights can be a challenge, given their complexity; for example, the cancer genome atlas contains multi-omic and clinical profiles of 11K+ patients across 33 cancer subtypes, over 500K files and 1 billion measurements. This talk will outline (1) how we are building the infrastructure and methods that couple data-driven approaches with a knowledgebase to enable mechanistic modeling of public data sources, (2) the challenges we face when modeling these kinds of data e.g., overfitting of models due to large feature sets on small sample sizes, and (3) a case study on one approach where we mined TCGA for mechanistic insights.

ORAL PRESENTATIONS

OP01: ABSTRACT WITHDRAWN

OP02: Enabling the next generation of microbiome science with QIIME 2
Presenting Author: Greg Caporaso, Northern Arizona University

OP03: Evolutionary Action is a unifying framework for assessing missense variant structures within and across phyla
Presenting Author: Nicholas Abel, Baylor College of Medicine

OP04: A toxicogenomics approach to identify liver and kidney injuries
Presenting Author: Patric Schyman, Biotechnology HPC Software Applications Institute (BHSAI)

OP05: Towards Automating Computational Phenotyping: Exploring the Trade-offs of Different Vocabulary Mapping Strategies
Presenting Author: Tiffany Callahan, University of Colorado Denver Anschutz Medical Campus- Computational Bioscience Program

OP06: Open PBTA: Collaborative analysis of the Pediatric Brain Tumor Atlas
Presenting Author: Joshua Shapiro, Childhood Cancer Data Lab (Alex's Lemonade Stand Foundation)

OP07: Data Discovery Engine: A web-based toolset for maximizing data discoverability and promoting reusable data-sharing best practices
Presenting Author: Marco Cano, Scripps Research

OP08: Hypergraph Analytics for Computational Virology
Presenting Author: Cliff Joslyn, Pacific Northwest National Laboratory

OP09: Containerized pipeline for the identification of compound heterozygous variants in trios
Presenting Author: Dustin Miller, Brigham Young University

OP10: The design of an interactive lung map for studying premalignant lesions in the lung over time
Presenting Author: Carsten Görg, Colorado School of Public Health

OP11: Apollo: an efficient tool to collaboratively refine and attribute genome-wide genomic annotations
Presenting Author: Nathan Dunn, University of California, Berkeley

OP12: An online end-to-end pipeline for virus phylogeography that leverages Natural Language Processing for finding host locations
Presenting Author: Matthew Scotch, Arizona State University

OP13: Evolutionary Action as a Tool for Quantifying Differentiation Across the Primate Family Tree
Presenting Author: Harley Peters, Baylor College of Medicine

OP14: Haplocravat: LD-based calculation built on top of a platform to annotate variants
Presenting Author: Ben Busby, Mountain Genomics/Johns Hopkins University

OP15: Characterizing the Regulatory Framework in an Aggressive Breast Cancer Phenotype: A Bayesian Regression-Based Enhancement
Presenting Author: George Acquaah-Mensah, Massachusetts college of Pharmacy & Health Sciences

OP16: Discovering Subclones in Tumors Sequenced at Standard Depths
Presenting Author: Li Liu, Arizona State University

OP17: Pathogenic Synonymous Variants Are More Likely to Affect Codon Usage Biases than Benign Synonymous Variants
Presenting Author: Justin Miller, Brigham Young University

OP18: Map and model — moving from observation to prediction in toxicogenomics
Presenting Author: Wibke Busch, Helmholtz Centre for Environmental Research - UFZ

OP19: Combining the Evolutionary Trace Algorithm and Covariation Metrics Yields Improved Structural Predictions
Presenting Author: Daniel Konecki, Baylor College of Medicine

OP20: Toxicological Mechanistic Inference: Generating mechanistic explanations of adverse outcomes
Presenting Author: Ignacio Tripodi, University of Colorado, Boulder

OP21: Giving credit where credit is due: How to make more meaningful connections between people and their roles, work, and impact
Presenting Author: Nicole Vasilevsky, Oregon Health & Science University

OP22: Nearest-neighbor Projected-Distance Regression to detect network interactions and control for confounders and multiple testing
Presenting Author: Trang Le, University of Pennsylvania

OP23: Deep learning enables in silico chemical-effect prediction
Presenting Author: Jana Schor, Helmholtz Centre for Environmental Research - UFZ

ORAL PRESENTATIONS

OP24: Correlations and curses of dimensionality: optimizing k in nearest-neighbor feature selection
Presenting Author: Bryan Dawkins, University of Tulsa

OP25: Deep Learning based Multi-view model for deciphering gene regulatory keywords
Presenting Author: Pramod Bharadwaj Chandrashekar, Arizona State University

OP26: Identifying optimal mouse models for human asthma using a novel modeling approach
Presenting Author: Yihsuan Tsai, University of North Carolina at Chapel Hill

OP27: Bridging the Bioinformatics Knowledge Gap in the Pediatric Cancer Research Community with the Childhood Cancer Data Lab workshops
Presenting Author: Chante Bethell, Childhood Cancer Data Lab (Alex's Lemonade Stand Foundation)

OP28: Hetnet connectivity search provides rapid insights into how two biomedical entities are related
Presenting Author: Daniel Himmelstein, University of Pennsylvania

OP29: Landmark and Cancer-Relevant Gene Selection of RNA Sequencing Data for Survival Analysis
Presenting Author: Carly Clayman, Penn State University - Great Valley

OP30: The impact of undesired technical variability on large-scale data compendia
Presenting Author: Alexandra Lee, Genomics and Computational Biology Graduate Program, University of Pennsylvania, Philadelphia, PA, USA; Department of Systems Pharmacology and Translational Therapeutics, University of Pennsylvania,

OP31: A pedigree-level examination of Schistosoma japonicum following schistosomiasis reemergence in rural China
Presenting Author: Laura Timm, University of Colorado - Anschutz Medical Campus

OP32: high resolution proteomics and genomics of CNDP1 repeat variants linked to diabetic nephropathy
Presenting Author: Nicholas Kinney, Edward Via College of Osteopathic Medicine

OP33: Computational Analysis of Kinesin Mutations Implicated in Hereditary Spastic Paraplegias
Presenting Author: Shaolei Teng, Howard University

OP34: Enabling structure-based data-driven selection of targets for cancer immunotherapy
Presenting Author: Dinler Antunes, Rice University

OP35: Biotherapeutic Protein Immunogenicity Risk Assessment with TCPro
Presenting Author: Osman Yogurtcu, FDA

OP36: Large-scale phylogenetic analysis reveals different sequence divergence patterns in orthologous and paralogous proteins
Presenting Author: Joseph Ahrens, University of Colorado Denver Anschutz Medical Campus

OP37: A new framework for clustering single cell RNA-seq data
Presenting Author: Ziyou Ren, Northwestern University

OP38: How to annotate what we don't know
Presenting Author: Mayla Boguslav, University of Colorado Anschutz Medical Campus

OP39: Filtering, classification, and selection of new knowledge for model assembly and extension
Presenting Author: Natasa Miskov-Zivanov, University of Pittsburgh

OP40: Pathway-based Single-Cell RNA-Seq Classification, Clustering, and Construction of Gene-Gene Interactions Networks Using Random Forests
Presenting Author: Herbert Pang, University of Hong Kong

POSTER PRESENTATIONS

LOCATION

Viceroy Hotel Ballroom

POSTER SESSION HOURS

The Poster Session with authors present will be on Friday evening. Poster Presenters must be available for presentation during the scheduled poster session.

POSTER NUMBER ASSIGNMENTS

Posters will be assigned even and odd numbers for presentation times. Please put your poster on the poster board corresponding to the number assigned. (The poster board dimensions are 4 feet high x 4 feet wide. Tacks will be provided for securing your poster to the board.)

SCHEDULE

FRIDAY, DECEMBER 6	**12:00 PM – 6:00 PM**	**SET UP POSTERS**
	6:30 PM – 7:30 PM	**POSTER SESSION WITH AUTHORS** *Even Number Posters*
	7:30 PM – 8:30 PM	**POSTER SESSION WITH AUTHORS** *Odd Number Posters**

*** Authors please remove posters from boards at end of this session.***

POSTER SESSION LIST

P01: Evolutionary Action is a unifying framework for assessing missense variant structures within and across phyla
Presenting Author: Nicholas Abel, Baylor College of Medicine

P02: Discovering Subclones in Tumors Sequenced at Standard Depths
Presenting Author: Navid Ahmadinejad, Arizona State University

P03: Meta-analysis and Machine Learning Classification for Dilated Cardiomyopathy Using Cardiac Transcriptomics Data
Presenting Author: Ahmad Alimadadi, Program in Bioinformatics and Department of Physiology and Pharmacology, University of Toledo College of Medicine and Life Sciences

P04: A web application for annotating tabular data with terms from biomedical ontologies
Presenting Author: Elizabeth Anderson, Brigham Young University

P05: Enabling structure-based data-driven selection of targets for cancer immunotherapy
Presenting Author: Dinler Antunes, Rice University

P06: ABSTRACT WITHDRAWN

P07: ABSTRACT WITHDRAWN

P08: Pathway-Regularized Matrix Factorization
Presenting Author: Aaron Baker, University of Wisconsin-Madison

P09: The Use of Machine Learning for Modeling a Clinical Decision Support for Predicting Postpartum Depression
Presenting Author: Houda Benlhabib, University of Washington, Biomedical Informatics and Medical Education

P10: Bridging the Bioinformatics Knowledge Gap in the Pediatric Cancer Research Community with the Childhood Cancer Data Lab workshops
Presenting Author: Chante Bethell, Childhood Cancer Data Lab (Alex's Lemonade Stand Foundation)

P11: Map and model — moving from observation to prediction in toxicogenomics
Presenting Author: Wibke Busch, Helmholtz Centre for Environmental Research - UFZ

P12: Data Discovery Engine: A web-based toolset for maximizing data discoverability and promoting reusable data-sharing best practices
Presenting Author: Marco Cano, Scripps Research

POSTER PRESENTATIONS

P13: A Machine Learning tool can assign function to phage proteins
Presenting Author: Vito Cantu, San Diego State University

P14: Identification of a lead compound for selective inhibition of Nav1.7 to treat chronic pain
Presenting Author: Sharat Chandra, Duke University

P15: Deep Learning based Multi-view model for deciphering gene regulatory keywords
Presenting Author: Pramod Bharadwaj Chandrashekar, Arizona State University

P16: Landmark and Cancer-Relevant Gene Selection of RNA Sequencing Data for Survival Analysis
Presenting Author: Carly Clayman, Penn State University - Great Valley

P17: Polygenic Risk Score Knowledge Base: A Web-based Application for Calculating Polygenic Risk Scores
Presenting Author: Matthew Cloward, Brigham Young University

P18: A comprehensive analysis of orthologous genes across all domains
Presenting Author: Lauren Cutler, Brigham Young University

P19: Using Mendelian Randomization to Assess Disease Causality
Presenting Author: Louisa Dayton, Brigham Young University

P20: ABSTRACT WITHDRAWN

P21: A Novel One-Class classification Approach to Accurately Predict Disease-Gene Association
Presenting Author: Abdollah Dehzangi, Morgan State University

P22: Pipelines, Workflows and Virtualization to Build Institutional Informatics Capacity
Presenting Author: Aaron Dickey, Unites States Department of Agriculture

P23: Transcriptional changes observed in a mouse contained infection model of TB identify human LTB+ individuals at low risk of progression to active disease.
Presenting Author: Fergal Duffy, Seattle Children's Research Institute

P24: Apollo: an efficient tool to collaboratively refine and attribute genome-wide genomic annotations
Presenting Author: Nathan Dunn, University of California, Berkeley

P25: Global phylogeography and ancient evolution of the widespread human gut virus crAssphage
Presenting Author: Robert Edwards, San Diego State University

P26: Urine pellet heterogeneity requires meticulous balancing of RNAseq libraries
Presenting Author: Felix Eichinger, University of Michigan

P27: Using Mutual Information to Validate Functional Interactions Between Clusterin and Amyloid Precursor Protein
Presenting Author: Austin Gale, Brigham Young University

P28: Computational Discovery of Novel Phages in the Human Gut Metagenome
Presenting Author: Melissa Giluso, San Diego State University

P29: Benchmarking Viral Identification Tools in Complex Simulated Metagenomes
Presenting Author: Cody Glickman, University of Colorado Anschutz

P30: The design of an interactive lung map for studying premalignant lesions in the lung over time
Presenting Author: Carsten Görg, Colorado School of Public Health

P31: pipelineTools: RStudio based NGS workflow, reporting and teaching system
Presenting Author: Graham Hamilton, University of Glasgow

P32: Hetnet connectivity search provides rapid insights into how two biomedical entities are related
Presenting Author: Daniel Himmelstein, University of Pennsylvania

P33: CUBAP: An Interactive Web Portal for Analyzing Codon Usage Bias Across Populations
Presenting Author: Matthew Hodgman, Brigham Young University

P34: Food Preservatives Induce Proteobacteria Dysbiosis in Human-Microbiota Associated Nod2-deficient Mice
Presenting Author: Tomas Hrncir, The Czech Academy of Sciences

P35: Predicting Clinical Dementia Rating Using Blood RNA Levels
Presenting Author: Erik Huckvale, Brigham Young University

P36: Polly Discover: Integration of public omics data and metadata using machine learning to enable asset discovery
Presenting Author: Brian Dranka, Elucidata

P37: Hypergraph Analytics for Computational Virology
Presenting Author: Cliff Joslyn, Pacific Northwest National Laboratory

POSTER PRESENTATIONS

P38: Analyzing the Signature of GPCR conformational Changes
Presenting Author: rafeed khleif, California State University Northridge

P39: High resolution proteomics and genomics of CNDP1 repeat variants linked to diabetic nephropathy
Presenting Author: Nicholas Kinney, Edward Via College of Osteopathic Medicine

P40: Combining the Evolutionary Trace Algorithm and Covariation Metrics Yields Improved Structural Predictions
Presenting Author: Daniel Konecki, Baylor College of Medicine

P41: ABSTRACT WITHDRAWN

P42: Expanding polygenic risk scores to include automatic genotype encodings and gene-gene interactions
Presenting Author: Trang Le, University of Pennsylvania

P43: The impact of undesired technical variability on large-scale data compendia
Presenting Author: Alexandra Lee, Genomics and Computational Biology Graduate Program, University of Pennsylvania, Philadelphia, PA, USA; Department of Systems Pharmacology and Translational Therapeutics, University of Pennsylvania,

P44: ABSTRACT WITHDRAWN

P45: PyFBA: from Genomics to Metabolomics
Presenting Author: Shane Levi, San Diego State University

P46: ABSTRACT WITHDRAWN

P47: Effective Targeted Drug Prediction for Cancer Based on Genetic Mutations
Presenting Author: Darsh Mandera, Jesuit High School

P48: Identifying Protein-Metabolite Networks associated with COPD Phenotypes
Presenting Author: Emily Mastej, University of Colorado Anschutz Medical Campus

P49: Identification of cell signaling pathways based on biochemical reaction kinetics repositories
Presenting Author: Gustavo Matos, University of Sao Paulo

P50: Using Ramp Sequences to Identify Causes of Disease Association
Presenting Author: Lauren McKinnon, Brigham Young University

P51: Predicting progression-free interval for cancer patients based on heterogeneous combinations of high-throughput molecular data
Presenting Author: Nathan Mella, Brigham Young University

P52: Codon Usage Biases Have Significant Implications in Population Stratification
Presenting Author: Taylor Meurs, Brigham Young University

P53: Containerized pipeline for the identification of compound heterozygous variants in trios
Presenting Author: Dustin Miller, Brigham Young University

P54: Pathogenic Synonymous Variants Are More Likely to Affect Codon Usage Biases than Benign Synonymous Variants
Presenting Author: Justin Miller, Brigham Young University

P55: Filtering, classification, and selection of new knowledge for model assembly and extension
Presenting Author: Natasa Miskov-Zivanov, University of Pittsburgh

P56: Hic-Pipeline: A Kepler And Spark-Based Scalable Workflow For Normalized Contact Map Creation
Presenting Author: Sanjay Nagaraj, Baylor College Of Medicine

P57: ABSTRACT WITHDRAWN

P58: Robust discovery of causal gene networks via measurement error estimation and correction
Presenting Author: Manikandan Narayanan, Indian Institute of Technology Madras

P59: Reusing label functions to extract multiple types of relationships from biomedical abstracts at scale
Presenting Author: David Nicholson, University of Pennsylvania

P60: Model selection for clinical metabolomics: comparing the power of different optimization approaches for coronary artery disease diagnosis prediction.
Presenting Author: Alena Orlenko, University of Pennsylvania

P61: The Utility of Polygenic Risk Scores in High-risk Pedigrees
Presenting Author: Madeline Page, Brigham Young University

P62: ABSTRACT WITHDRAWN

P63: Medication Profiling through Tensor Factorization: A case study on commercial pharmacy claims
Presenting Author: Yubin Park, WithMe Health, Inc.

P64: Evolutionary Action as a Tool for Quantifying Differentiation Across the Primate Family Tree
Presenting Author: Harley Peters, Baylor College of Medicine

P65: Assessing long-read correction and polishing strategies for genome assembly
Presenting Author: Brandon Pickett, Brigham Young University

POSTER PRESENTATIONS

P66: ABSTRACT WITHDRAWN

P67: Different Metabolic RNA Levels Exist in Alzheimer's Disease Brains
Presenting Author: Karl Ringger, Brigham Young University

P68: Effect of an axially oriented electric dipole moment on the amyloid-ß(25-35) aggregation and cytotoxicity
Presenting Author: Eduardo Romero, University of Central Florida

P69: Deep learning enables in silico chemical-effect prediction
Presenting Author: Jana Schor, Helmholtz Centre for Environmental Research - UFZ

P70: An online end-to-end pipeline for virus phylogeography that leverages Natural Language Processing for finding host locations
Presenting Author: Matthew Scotch, Arizona State University

P71: Open PBTA: Collaborative analysis of the Pediatric Brain Tumor Atlas
Presenting Author: Joshua Shapiro, Childhood Cancer Data Lab (Alex's Lemonade Stand Foundation)

P72: Technical Bias Correction of Sequencing Libraries Using Wavelet Transform Analysis and Clustering
Presenting Author: Rutendo Sigauke, Univeristy of Colorado

P73: geneHarmony: An Interactive Web Application that Automates the Manual Process of Bacteriophage Genome Annotation
Presenting Author: Erica Suh, Brigham Young University

P74: Natural Language Processing (NLP) in FDA Adverse Event Reporting System (FAERS) to Improve Extraction of Rare, Severe Adverse Drug Event-Product Pairs from Unstructured Data: Stevens - Johnson Syndrome and Toxic Epidermal Necrolysis (SJS/TEN) as an E
Presenting Author: Katherine Sullivan, University of Colorado, Anschutz Medical Campus

P75: MetaPro: A scalable and reproducible data processing and analysis pipeline for metatranscriptomic investigation of microbial communities
Presenting Author: Billy Taj, Hospital for Sick Children

P76: A review on multiple sequence alignment algorithms
Presenting Author: Alice Tan, White Oaks Secondary School

P77: Computational Analysis of Kinesin Mutations Implicated in Hereditary Spastic Paraplegias
Presenting Author: Shaolei Teng, Howard University

P78: A pan-cancer 3-gene signature to predict dormancy
Presenting Author: Ivy Tran, Rutgers University - Camden

P79: Identifying optimal mouse models for human asthma using a novel modeling approach
Presenting Author: Yihsuan Tsai, University of North Carolina at Chapel Hill

P80: Mitochondrial DNA Deletions and Copy Number in Whole Genome Sequencing (WGS) Data: Analyses of Aging and Parkinson's Disease using Brain and Blood
Presenting Author: David Tyrpak, University of Southern California

P81: ABSTRACT WITHDRAWN

P82: Mondo Disease Ontology: harmonizing disease concepts around the world
Presenting Author: Nicole Vasilevsky, Oregon Health & Science University

P83: Giving credit where credit is due: How to make more meaningful connections between people and their roles, work, and impact
Presenting Author: Nicole Vasilevsky, Oregon Health & Science University

P84: Leveraging Familial-based Relationships for Rare Variant Discovery
Presenting Author: Liz Ward, Brigham Young University

P85: Identifying Functional Relationships Using Protein Coevolution
Presenting Author: Katrisa Ward, Brigham Young University

P86: ABSTRACT WITHDRAWN*e*

P87: Optimizing storage and querying of massive biological datasets of a tabular nature
Presenting Author: James Wengler, Brigham Young University

P88: Biotherapeutic Protein Immunogenicity Risk Assessment with TCPro
Presenting Author: Osman Yogurtcu, FDA

P89: ABSTRACT WITHDRAWN

P90: Structuring and crawling distributed biomedical metadata using schema.org standard
Presenting Author: Xinghua Zhou, The Scripps Research Institute