ACM-BCB’18
Proceedings of the 2018 ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics

Sponsored by:
ACM SIGBio

Supported by:
National Science Foundation
# ACM-BCB Organization Committee

## General Chairs
- **Amarda Shehu**  
  George Mason University, USA  
- **Cathy H. Wu**  
  University of Delaware, USA  

## Program Chairs
- **Christina Boucher**  
  University of Florida, USA  
- **Jing Li**  
  Case Western Reserve University, USA  
- **Hongfang Liu**  
  Mayo Clinic, USA  
- **Mihai Pop**  
  University of Maryland, USA  

## Workshop Chairs
- **Mario Cannataro**  
  University Magna Graecia of Catanzaro, Italy  
- **Ka Yee Yeung**  
  University of Washington Tacoma, USA  

## Tutorial Chairs
- **Rezarta Islamaj Dogan, NCBI, NLM**  
  National Institutes of Health, USA  
- **Casey Overby Taylor**  
  Johns Hopkins University, USA  

## Poster Chairs
- **Sheida Navabi**  
  University of Connecticut, USA  

## Women in Bioinformatics (WIB) Chair
- **May D. Wang**  
  Georgia Institute of Technology & Emory University, USA  

## Demo and Exhibit Chair
- **Nikos Kyrpides**  
  DOE Joint Genome Institute, USA  

## Highlights Chairs
- **Lenore Cowen**  
  Tufts University, USA  
- **Xiaokqian Jiang**  
  University of Texas Health Science Center at Houston, USA  

## Registration Chair
- **Preetam Ghosh**  
  Virginia Commonwealth University  

## Student Activities Chairs
- **Nasrin Akhter**  
  George Mason University, USA  
- **Riza Bautista**  
  University of Delaware, USA  

## Local Arrangement Chairs
- **Amarda Shehu**  
  George Mason University  
- **Liliana Florea**  
  Johns Hopkins University  

## Student Travel Award Chair
- **May D. Wang**  
  Georgia Institute of Technology & Emory University, USA  

## Proceedings Chairs
- **Yang Shen**  
  Texas A&M University, USA  
- **Xinghua Mindy Shi**  
  University of North Carolina at Charlotte, USA  

## Publicity Chairs
- **Debswapna Bhattacharya**  
  Auburn University, USA  
- **A. Ercument Cicek**  
  Bilkent University, Turkey  
- **Oznur Tastan**  
  Sabanci University, Turkey  
- **Pierangelo Veltri**  
  University Magna Graecia of Catanzaro, Italy  

## Web Admins
- **Aman Sawhney**  
  University of Delaware, USA  
- **Sachin Gavali**  
  University of Delaware, USA  

WiFi Access Information

Network: JWMarriott_CONFERENCE   Password: ACMBCB2018

Conference Schedule

<table>
<thead>
<tr>
<th>Wednesday Aug. 29</th>
<th>Thursday Aug. 30</th>
<th>Friday Aug. 31</th>
<th>Saturday Sep. 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>8:00am – 8:30am</td>
<td>8:00am – 8:30am</td>
<td>8:00am – 8:30am</td>
<td>8:00am – 8:30am</td>
</tr>
<tr>
<td>Light Breakfast</td>
<td>Light Breakfast</td>
<td>Light Breakfast</td>
<td>Light Breakfast</td>
</tr>
<tr>
<td>8:45am – 9:00am</td>
<td>8:45am – 9:00am</td>
<td>8:45am – 9:00am</td>
<td>8:45am – 9:00am</td>
</tr>
<tr>
<td>Opening Remarks</td>
<td>Opening Remarks</td>
<td>Opening Remarks</td>
<td>Opening Remarks</td>
</tr>
<tr>
<td>9:00am – 10:00am</td>
<td>9:00am – 10:00am</td>
<td>9:00am – 10:00am</td>
<td>9:00am – 10:00am</td>
</tr>
<tr>
<td>Keynote Talk 1</td>
<td>Keynote Talk 2</td>
<td>Keynote Talk 3</td>
<td>Keynote Talk 3</td>
</tr>
<tr>
<td>10:00am – 10:30am</td>
<td>10:00am – 10:30am</td>
<td>10:00am – 10:30am</td>
<td>10:00am – 10:30am</td>
</tr>
<tr>
<td>Break</td>
<td>Break</td>
<td>Break</td>
<td>Break</td>
</tr>
<tr>
<td>10:30am – Noon</td>
<td>10:30am – Noon</td>
<td>10:30am – Noon</td>
<td>10:30am – Noon</td>
</tr>
<tr>
<td>Paper Session 1</td>
<td>Paper Session 3</td>
<td>Paper Session 5</td>
<td>Paper Session 5</td>
</tr>
<tr>
<td>Noon – 2:00pm</td>
<td>Noon – 2:00pm</td>
<td>12:15pm – 1:45pm</td>
<td></td>
</tr>
<tr>
<td>Lunch at</td>
<td>Lunch at</td>
<td>Lunch at</td>
<td></td>
</tr>
<tr>
<td>Penn Avenue Terrace</td>
<td>Penn Avenue Terrace</td>
<td>Penn Avenue Terrace</td>
<td></td>
</tr>
<tr>
<td>&amp; Funding Panel</td>
<td>&amp; Funding Panel</td>
<td>&amp; Funding Panel</td>
<td>&amp; Funding Panel</td>
</tr>
<tr>
<td>1:00pm – 3:30pm</td>
<td>2:00pm – 3:30pm</td>
<td>1:45pm – 2:00pm</td>
<td></td>
</tr>
<tr>
<td>Grant Writing Workshop</td>
<td>Paper Session 4 &amp;</td>
<td>Closing Remarks</td>
<td></td>
</tr>
<tr>
<td></td>
<td>IEEE/ACM TCBB Editorial Board Meeting</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2:00pm – 3:30pm</td>
<td>Paper Session 2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Paper Session 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3:30pm – 4:00pm</td>
<td>3:30pm – 4:00pm</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Break</td>
<td>Break</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4:00pm – 5:30pm</td>
<td>4:00pm – 6:00pm</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ACM SIG-BIO Community Meeting</td>
<td>Student Research Forum</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6:00pm – 8:30pm</td>
<td>6:00pm – 8:00pm</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reception &amp; Poster Session</td>
<td>Dinner Banquet</td>
<td>(Pre-registration Required)</td>
<td></td>
</tr>
</tbody>
</table>
Workshops

Workshop W1

8:00am – 5:00pm
Location: Salon H-J

The Fifth International Workshop on Computational Network Biology: Modeling, Analysis, and Control (CNB-MAC 2018)

Dr. Byung-Jun Yoon and Dr. Xiaoning Qian, Texas A&M University, Dept. Electrical & Computer Engineering;
Dr. Tamer Kahveci, University of Florida, Dept. Computer and Information Science and Engineering;
Dr. Ranadip Pal, Texas Tech University, Electrical and Computer Engineering

Abstract: Next-generation high-throughput profiling technologies have enabled more systematic and comprehensive studies of living systems. Network models play crucial roles in understanding the complex interactions that govern biological systems, and their interactions with external environment. The inference and analysis of such complex networks and network-based analysis of large-scale measurement data have already shown strong potential for unveiling the key mechanisms of complex diseases as well as for designing improved therapeutic strategies. At the same time, the inference and analysis of complex biological networks pose new exciting challenges for computer science, signal processing, control, and statistics. We propose to organize the Fifth International Workshop on Computational Network Biology: Modeling, Analysis, and Control (CNB-MAC 2018) in conjunction with ACM-BCB 2018. The previous CNB-MAC workshops have been successfully held in conjunction with ACM-BCB 2014, ACM-BCB 2015, ACM-BCB 2016, ACM-BCB 2017, attracting a fair number of researchers interested in computational network biology.

The workshop aims to provide an international scientific forum for presenting recent advances in computational network biology that involve modeling, analysis, and control of biological systems under different conditions, and system-oriented analysis of large-scale OMICS data. The proposed full-day workshop will solicit (i) highlights that present advances in the field that have been reported in recent journal publications, (ii) extended abstracts for poster presentation at the workshop, which will provide an excellent venue for quick dissemination of the latest research results in computational network biology, and (iii) original research papers that report new research findings that have not been published elsewhere. Full length original research papers accepted for presentation at the workshop will be published in a supplement issue in partner journals that will be identified after the workshop proposal is accepted. The first and the second CNB-MAC workshops have partnered with EURASIP Journal on Bioinformatics and Systems Biology, and the third CNB-MAC workshop partnered with BMC Bioinformatics, BMC Systems Biology, and BMC Genomics. The fourth CNB-MAC workshop partnered with BMC Bioinformatics, BMC Systems Biology, BMC Genomics, and IET Systems Biology. The main emphasis
of the proposed workshop will be on rigorous mathematical or computational approaches in studying biological networks, analyzing large-scale OMICS data, and investigating mathematical models for human-microbiome-environment interactions.

**Workshop W2**

8:00am – 5:00pm  
Location: Salon A-B  
Computational Structural Bioinformatics Workshop (CSBW)  
Filip Jagodzinski, Dept. of Computer Science, Western Washington University; Brian Chen, Dept. of Computer Science and Engineering, Lehigh University; Kevin Molloy, Dept. of Computer Science, James Madison University

**Abstract:** The unique nature of protein and nucleotide structures has presented many computational challenges and opportunities. The fast accumulation and use of various sources of data has enabled a variety of novel analysis approaches and computational techniques, yielding a variety of insights and increased understanding of different biophysical phenomena and processes. This workshop aims to provide a dissemination and discussion forum for computational approaches related to protein structural discovery and analysis. We have successfully hosted the CSBW with BIBM in 2007-2009, 2011, 2012, and 2015, and with ACM-BCB in 2013, 2014, 2016 and 2017. Past workshops have been well attended, with approximately 12 oral presentations for peer reviewed submissions. In past years, we have also held a poster session during the workshop, to complement the conference poster session, and to encourage participation by undergraduate and graduate students in presenting & discussing on-going projects.

**Workshop W3**

8:00am – 5:00pm  
Location: Salon E  
7th Workshop on Computational Advances in Molecular Epidemiology (CAME 2018)  
Yury Khudyakov, Centers for Disease Control and Prevention; Ion Mandoiu, University of Connecticut; Pavel Skums and Alex Zelikovsky, Georgia State University

**Abstract:** The CAME workshop provides a forum for presentation and discussion of the latest computational research in molecular epidemiology. This multidisciplinary workshop will bring together field practitioners of molecular epidemiology, molecular evolutionists, population geneticists, medical researchers, bioinformaticians, statisticians and computer scientists interested in the latest developments in algorithms, mining, visualization, modeling, simulation and other methods of computational, statistical and mathematical analysis of genetic and molecular data in the epidemiological context.

Molecular epidemiology is essentially an integrative scientific discipline that considers molecular biological processes in specific epidemiological settings. It relates molecular biological events to etiology, distribution and prevention of disease in human
populations. Over years, molecular epidemiology became extensively fused with mathematical and computational science and immensely benefited from this tight association. The workshop will review the latest advancements in application of mathematical and computational approaches to molecular epidemiology.

**Workshop W4**

8:00am – Noon  
**Location: Russell/Hart/Cannon**  
**7th Parallel and Cloud-based Bioinformatics and Biomedicine (ParBio)**  
Giuseppe Agapito, Department of Medical and Surgical Sciences, University Magna Græcia of Catanzaro, Italy; Wes Lloyd, Institute of Technology, University of Washington - Tacoma

**Abstract:** Due to the availability of high-throughput platforms (e.g., next-generation sequencing, microarray and mass spectrometry) and clinical diagnostic tools (e.g., medical imaging), a recent trend in Bioinformatics and Biomedicine is the ever-increasing production of experimental and clinical data.

Considering the complex analysis pipelines often used in biomedical research, the bottleneck increasingly involves the storage, integration, and analysis of experimental data, as well as their correlation and integration with publicly available data banks.

Grid infrastructures may offer the huge data storage needed to store experimental and biomedical data, while parallel computing can be used for basic pre-processing (e.g., parallel BLAST, mpiBLAST) and for more advanced analysis (e.g. parallel data mining). In such a scenario, novel parallel architectures (e.g., CELL processors, GPUs, FPGA, hybrid CPU/FPGA) coupled with emerging programming models may overcome the limits posed by conventional computers to the mining and exploration of large amounts of data.

On the other hand, these technologies yet require great investments by biomedical and clinical institutions and are based on a traditional model where users often need to be aware and face different management problems, such as hardware and software management, data storage, software ownership, and prohibitive costs (different professional-level applications in the biomedical domain have a high starting cost that prevent many small laboratories to use them).

While parallel computing and Grid computing offer computational power and storage to address the overwhelming availability of data, Cloud Computing additionally hides the complexity of computing infrastructures, while also reducing the cost of data analysis tasks, demonstrating potential to transform the overall model of biomedical research and health related data science.

Cloud Computing, that offers scalable costs, increased accessibility, availability, and ease of application use while enabling potential collaboration among scientists, is already changing the computing business models in different sectors, and recently has been adopted to support bioinformatics (see for instance the recent JCVI Cloud Bio-Linux initiative) and biomedical domains. However, many problems remain to be solved, such as availability and safety of the data, privacy-related issues, availability of software platforms for rapid deployment, and the execution and billing of biomedical applications.
**Workshop W5**

1:00pm – 5:00pm  
Location: Russell/Hart/Cannon  
NCI Cloud Resources  
KanakaDurga Addepalli and Hsinyi (Steve) Tsang, National Cancer Institute

**Abstract:** Technological advancements have given us the ability to sequence genomes in great depths, and, consequently, generated an exponential growth in data. National Cancer Institute Cloud Resources (NCICR), formerly the NCI Cancer Genomics Cloud Pilots, were developed with a goal of enhancing the utility of cancer genomic data and facilitating analysis by co-locating cloud computing and petabyte-scale data. Based on commercial cloud architectures, the Cloud Resources offer the flexibility for users to utilize tools in the form of Docker containers, and tools can be joined to form complex workflows described by Common Workflow Language (CWL) or Workflow Description Language (WDL). The utility and application of the Cloud Resources has been expanded from cancer genomics to include microbe analysis, proteomics, imaging and other “omics” in the future. The cloud environment has proven to be a cost-effective, reproducible, reusable, interoperable, and user-friendly alternative to high-performance computing, with minimal overhead and setup requirements. These production-ready and highly scalable platforms represent a necessary step in a publicly available toolset meant to support open and Findable, Accessible, Interoperable, Reusable (FAIR) scientific research.

Through this demonstration workshop, participants will have the opportunity to

1. learn about the basic features of the NCICR  
2. create interoperable, containerized tools, and  
3. run genomic analysis on the three cloud-based platforms.

**Satellite Meeting S1**

8:00am – Noon  
Location: Salon F  
BioCreative/OHNLP Challenge 2018  
Majid Rastegar-Mojarad, Sijia Liu, Yanshan Wang, Naveed Afzal, Liwei Wang, Feichen Shen, Sunyang Fu, and Hongfang Liu, Department of Health Sciences Research, Mayo Clinic

**Abstract:** The application of Natural Language Processing (NLP) methods and resources to clinical and biomedical text has received growing attention over the past years, but progress has been limited by difficulties to access shared tools and resources, partially caused by patient privacy and data confidentiality constraints. Efforts to increase sharing and interoperability of the few existing resources are needed to facilitate the progress observed in the general NLP domain. Leveraging our research in corpus analysis and de-identification research, we have created multiple synthetic data sets for a couple of NLP tasks based on real clinical sentences. We
are organizing a challenge workshop to promote community efforts towards the advancement in clinical NLP. The challenge workshop will have two tasks:

- Task 1. Family History Information Extraction - contact Majid Rastegar-Mojarad (mojarad.majid@mayo.edu) or Sijia Liu (liu.sijia@mayo.edu)
- Task 2. Clinical Semantic Textual Similarity - contact Yanshan Wang (wang.yanshan@mayo.edu)

Tutorials

**Tutorial T1**

8:00am – 10:00am  
Location: Justice Room  
**Modeling Macromolecular Structures and Motions: Computational Methods for Sampling and Analysis of Energy Landscapes**  
Kevin Molloy, Nasrin Akhter, and Amarda Shehu,  
Department of Computer Science, George Mason University

**Abstract:** With biomolecular structure recognized as central to understanding mechanisms in the cell, computational chemists and biophysicists have spent significant efforts on modeling and analyzing structure and dynamics. While significant advances have been made, particularly in the design of sophisticated energetic models and molecular representations, such efforts are experiencing diminishing returns. One of the culprits is the low exploration capability of Molecular Dynamics- and Monte Carlo-based exploration algorithms. The impasse has attracted AI researchers bringing complementary tools, such as randomized search and stochastic optimization. The objective of this tutorial is three-fold. First, the tutorial will introduce students and researchers that attend ACM-BCB to stochastic optimization treatments and methodologies for understanding and elucidating the role of structure and dynamics in the function of biomolecules. Second, the tutorial will allow attendees to connect between structures, motions, and function via analysis tools that take an energy landscape view of the relationship between biomolecular structure, dynamics, and function. Third, the presentation will be enhanced via open-source software that permit hands-on exercises. One such software is developed in the Shehu Computational Biology laboratory and allows researchers both to integrate themselves in a new research domain as well as drive further research via plug-and-play capabilities. The hands-on approach in the tutorial will be beneficial to students and senior researchers keen to make their own contributions.
Tutorial T2

8:00am – 10:00am
Location: Salon C
Making Deep Learning Understandable for Analyzing Sequential Data about Gene Regulation
Yanjun (Jane) Qi, Ph.D., Department of Computer Science, School of Engineering and Applied Science, University of Virginia

Abstract: The past decade has seen a revolution in genomic technologies that enable a flood of genome-wide profiling of molecular elements on human genomes across many different tissue types. This massive-scale molecular data provides researchers with an unprecedented opportunity to understand gene regulation that can enable new insights into principles of life, the study of diseases, and the development of treatments and drugs. Computational challenges are the major bottlenecks for comprehensive genome-wide data analysis of gene regulation. Such data sets are complex, often ill-understood and at an unprecedented scale of data growth. Problems of this nature may be particularly well suited to deep learning techniques that recently show impressive results across a variety of domains. This tutorial aims to provide an extensive literature review about the state-of-the-art techniques in deep Learning, to examine how deep learning is enabling changes at analyzing datasets about gene regulations, and to foresee the potential of deep to transform several areas of biology and medicine.

Tutorial T3

10:00am – Noon
Location: Salon K
Interpretable Machine Learning in Healthcare

Abstract: This tutorial extensively covers the definitions, nuances, challenges, and requirements for the design of interpretable and explainable machine learning models and systems in healthcare. We discuss many uses in which interpretable machine learning models are needed in healthcare and how they should be deployed. Additionally, we explore the landscape of recent advances to address the challenges model interpretability in healthcare and also describe how one would go about choosing the right interpretable machine learning algorithm for a given problem in healthcare.
Tutorial T4

10:30am – Noon
Location: Salon C
Using BioDepot-workflow-Builder to Create and Execute Reproducible Bioinformatics Workflows
K. Y. Yeung, L.-H. Hung, and W. Lloyd,
Institute of Technology, University of Washington, Tacoma, WA, USA

Abstract: Reproducibility is essential for the verification and advancement of scientific research. It is often necessary, not just to recreate the code, but also the software and hardware environment to reproduce results of computational analyses. Software containers like Docker, that distribute the entire computing environment are rapidly gaining popularity in bioinformatics. Docker not only allows for the reproducible deployment of bioinformatics workflows, but also facilitates mix-and-match of components from different workflows that have complex and possibly conflicting software requirements. However, configuration and deployment of Docker, a command-line tool, can be exceedingly challenging for biomedical researchers with limited training in programming and technical skills.

Tutorial T5

1:00pm – 5:00pm
Location: Justice Room
Analysis of Sequencing Data
T. Sofer, Program in Sleep Medicine Epidemiology, Brigham and Women’s Hospital, Harvard Medical School; M. Graff, University of North Carolina at Chapel Hill, Department of Epidemiology

Abstract: Large Whole Exome and Genome Sequence (WES, WGS) data are becoming available for the biomedical community through NHLBI and NHGRI studies such as TOPMed and CCDG. These studies collected more than WGS from more than 100,000 each, with individuals contributed from multiple, existing, epidemiological studies, representing genetically and environmentally diverse populations. Because the goal of WGS studies are to identify rare variants association with health outcomes, these heterogeneous data have to be analyzed together. This represent new challenges due to their inherent heterogeneity, that was not seen earlier, when genetic association studies analyzed diverse populations separately, and later combined them in meta-analysis.

In this workshop, we will introduce the challenges and pitfalls in analyzing large sequencing data sets, and analysis approaches that address these challenges. At the end of the workshop, attenders will be familiar with these challenges, with a few software packages and how they address them, and will know how to look at the scientific literature to form a specific plan to analyze their own data, based on their specific structure and goals.

Workshop url: https://scholar.harvard.edu/tsofer/analysis-sequencing-data-tutorial
Tutorial T6

1:00pm – 3:30pm
Location: Salon K

Agile Clinical Decision Support Development and Implementation

Mujeeb A. Basit, MD, MMSc, Vaishnavi Kannan, MS, and Duwayne L. Willett, MD, MS,
University of Texas Southwestern Medical Center, Dallas, Texas

Abstract: Designing effective Clinical Decision Support (CDS) tools in an Electronic Health Record (EHR) can prove challenging, due to complex real-world scenarios and newly-discovered requirements. Deploying new CDS tools shares much in common with new product development, where “agile” principles and practices consistently prove effective. Agile methods can thus prove helpful on CDS projects, including time-boxed “sprints” and lightweight requirements gathering with User Stories. Modeling CDS behavior promotes unambiguous shared understanding of desired behavior, but risks analysis paralysis: an Agile Modeling approach can foster effective rapid-cycle CDS design and optimization. The agile practice of automated testing for test-driven design and regression testing can be applied to CDS development using open-source tools. Ongoing monitoring of CDS behavior once released to production can identify anomalies and prompt rapid-cycle redesign to further enhance CDS effectiveness. The workshop participant will learn about these topics in interactive didactic sessions, with time for practicing the techniques taught.

Tutorial T7

1:00pm – 3:00pm
Location: Salon C

Rapidly identifying Disease-Associated Rare Variants using Annotation and Machine Learning at Whole-genome Scale Online

Alex Kotlar, Department of Human Genetics, Emory University; Thomas S. Wingo, Department of Neurology, Emory University, Division of Neurology, Atlanta VA Medical Center

Abstract: Accurately selecting disease-associated alleles from large sequencing experiments remains technically challenging. During this tutorial, participants will learn how to use a new variant annotation and classification tool called Bystro (https://bystro.io/) to perform rare-variant association studies at the scale of whole-genome and whole-exome experiments. Bystro is the first online, cloud-based application that makes variant annotation and filtering accessible to all researchers for terabyte-sized whole-genome experiments containing thousands of samples. Its key innovation is a general-purpose, natural-language search engine that enables users to identify and export alleles and samples of interest in milliseconds. Participants will be shown how to 1) annotate Variant Call Format (VCF) data online using Bystro, 2) remove low-quality samples based on transition/transversion, silent/replacement, theta, and other key quality control metrics, 3) use Bystro’s search engine to remove low quality alleles as well as sites unlikely to be associated with disease, 4) apply machine learning methods including support vector machines (SVM) and classification trees.
(LightGBM) to classify disease-associated alleles, 5) perform rare-variant association tests (R/SKAT) on the filtered and quality-controlled alleles to identify disease associated genes. Most steps will be performed online in Bystro.

**Student Social Event**

<table>
<thead>
<tr>
<th>Time</th>
<th>Activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>3:00pm – 3:15pm</td>
<td>Gather at Capitol Foyer</td>
</tr>
<tr>
<td>3:30pm – 5:30pm</td>
<td>Museum Visit</td>
</tr>
<tr>
<td>6:00pm – 7:30pm</td>
<td>Meet and Greet at Proper21</td>
</tr>
</tbody>
</table>

Chairs: Nasrin Akhter, George Mason University, USA; Riza Bautista, University of Delaware, USA
Thursday, August 30, 2018

8:00am – 8:30am Light Breakfast at Capitol Foyer

Opening and Welcome Remarks

8:45am – 9:00am
Location: Salon E, F, G
Session Chairs: Cathy Wu, University of Delaware and Amarda Shehu, George Mason University

Keynote Talk 1

9:00am – 10:00am
Location: Salon E, F, G
Sequence, Structure and Network Methods to Uncover Cancer Genes
Mona Singh, PhD, Princeton University
Session Chair: Amarda Shehu, George Mason University

Abstract: A major aim of cancer genomics is to pinpoint which somatically mutated genes are involved in tumor initiation and progression. This is a difficult task, as numerous somatic mutations are typically observed in each cancer genome, only a subset of which are cancer-relevant, and very few genes are found to be somatically mutated across large numbers of individuals. In this talk, I will overview three methods my group has introduced for identifying cancer genes. First, I will present a framework for uncovering cancer genes, differential mutation analysis, that compares the mutational profiles of genes across cancer genomes with their natural germline variation across healthy individuals. Next, I will show how to leverage per-individual mutational profiles within the context of protein-protein interaction networks in order to identify small connected subnetworks of genes that, while not individually frequently mutated, comprise pathways that are altered across (i.e., “cover”) a large fraction of individuals. Finally, I will demonstrate that cancer genes can be discovered by identifying genes whose interaction interfaces are enriched in somatic mutations. Overall, these methods recapitulate known cancer driver genes, and discover novel, and sometimes rarely-mutated, genes with likely roles in cancer.

Biography: Mona Singh obtained her AB and SM degrees at Harvard University, and her PhD at MIT, all three in Computer Science. She did postdoctoral work at the Whitehead Institute for Biomedical Research. She has been on the faculty at Princeton since 1999, and currently she is Professor of Computer Science in the Department of Computer Science and the Lewis-Sigler Institute for Integrative Genomics. Her group works broadly in computational molecular biology, as well as its interface with machine learning and algorithms. Her group is especially interested in developing methods for predicting and characterizing protein specificity,
interactions, and networks. She is currently a Methods Editor at PLOS Computational Biology, has been program committee chair for several major computational biology conferences, including ISMB (2010), WABI (2010), ACM-BCB (2012), and RECOMB (2016), and has been Chair of the NIH Modeling and Analysis of Biological Systems Study Section (2012-2014). She received the Presidential Early Career Award for Scientists and Engineers (PECASE) in 2001, and is an ISCB Fellow.

10:00am – 10:30am Morning Break

**Paper Session 1**

**10:30am – Noon**

**Session 1: Network Analysis**

*Location: Salon D*

**Session Chair:** Jing Li, Case Western Reserve University

1. **H: 10:30-10:55am**
   Aligning dynamic networks with DynaWAVE. Vijayan V, Milenkovic T.

2. **L: 10:55-11:20am**
   Target gene prediction of transcription factor using a new neighborhood-regularized tri-factorization one-class collaborative filtering algorithm. Hansaim Lim and Lei Xie.

3. **L: 11:20-11:45am**
   A Decomposition-based Approach towards the Control of Boolean Networks. Soumya Paul, Cui Su, Jun Pang and Andrzej Mizera.

**Session 2: Sequence Analysis**

*Location: Salon E*

**Session Chair:** Tamer Kahveci University of Florida

1. **H: 10:30-10:55am**
   ROP: dumpster diving in RNA-sequencing to find the source of 1 trillion reads across diverse adult human tissues. Serghei Mangul.

2. **L: 10:55-11:20am**
   Towards selective-alignment: Bridging the accuracy gap between alignment-based and alignment-free transcript quantification. Hirak Sarkar, Mohsen Zakeri, Laraib Malik and Rob Patro.

3. **L: 11:20-11:45am**
   ULTRA: A Model Based Tool to Detect Tandem Repeats. Daniel Olson and Travis Wheeler

**Session 3: Genetics and Evolution**

*Location: Salon F*

**Session Chair:** Hongfang Liu Mayo Clinic

1. **H: 10:30-10:55am**

2. **L: 10:55-11:20am**

3. **L: 11:20-11:45am**

**Session 4: Gene Expression Analysis**

*Location: Salon G*

**Session Chair:** Christina Boucher University of Florida

1. **L: 10:30-10:55am**

2. **L: 10:55-11:20am**
   Deep Learning Convolutional Neural Network for Cancer Diagnosis using Gene Expression Pattern. Nan Deng

3. **L: 11:20-11:45am**
   Deep Learning Based Tumor Type Classification Using Gene Expression Data. Boyu Lyu and Anamul Haque.
4. **S: 11:45am-Noon**

4. **S: 11:45am-Noon**
Measuring the mappability spectrum of reference genome assemblies. Zachary Stephens and Ravishankar Iyer.

4. **S: 11:45am-Noon**

4. **S: 11:45am-Noon**
A Distributed Constrained Non-negative Matrix Factorization Algorithm for Time-Series Gene Expression Dat. Matthew Dyer and Julian Dymacek.

---

**Noon – 2:00pm Lunch at Penn Avenue Terrace**

**Funding Agency Panel**

Noon – 2:00pm
Location: Russell/Hart/Cannon
Session Chairs: Cathy Wu, University of Delaware and Amarda Shehu, George Mason University

---

**Panelists**

**Mitra Basu**
Program Director, Division of Computing and Communication Foundations (CCF)
Directorate for Computer & Information Science & Engineering (CISE)
National Science Foundation (NSF)

**Wendy Nielsen**
Program Director, Division of Information and Intelligent Systems (IIS)
Directorate for Computer & Information Science & Engineering (CISE)
National Science Foundation (NSF)

**Jennifer Couch**
Chief, Structural Biology and Molecular Applications Branch
Division of Cancer Biology (DCB)
National Cancer Institute (NCI)

**Susan Gregurick**
Director, Division of Biophysics, Biomedical Technology, and Computational Biosciences (BBCB)
National Institute of General Medical Sciences (NIGMS)

**Daniel Drell**
Program Manager, Biological Systems Science Division (BSSD)
Office of Biological & Environmental Research (OBER)
Department of Energy (DOE)
Hands-on NIH Grant Writing Workshop

1:30pm – 3:30pm
Location: Salon G
Session Chairs: Cathy Wu, University of Delaware and Amarda Shehu, George Mason University

Speakers

Veerassamy “Ravi” Ravichandran
Program Director
Division of Biophysics, Biomedical Technology, and Computational Biology (BBCB)
National Institute of General Medical Sciences (NIGMS)
National Institutes of Health (NIH)

Wenchi Liang
Scientific Review Officer
Biodata Management and Analysis (BDMA) Study Section
Center for Scientific Review (CSR)
National Institutes of Health (NIH)

Paper Session 2

2:00pm – 3:30pm

Session 5: Medical Informatics
Location: Salon D
Session Chair: Naveen Ashish Fredhutch

Session 6: Sequence Analysis 2
Location: Salon E
Session Chair: Christina Boucher University of Florida

Session 7: Phenotypes and Diseases
Location: Salon F
Session Chair: Mukul Bansal University of Connecticut

1. H: 2:00-2:25pm
Co-occurrence of medical conditions: Exposing patterns through probabilistic topic modeling of snomed codes.
Bhattacharya M, Jurkowitz C, Shatkay H.

1. H: 2:00-2:25pm

1. H: 2:00-2:25pm
RegMIL: Phenotype Classification From Metagenomic Data. Mohammad Arifur Rahman and Huzefa Rangwala.
<table>
<thead>
<tr>
<th>Time</th>
<th>Session</th>
<th>Title</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>2:25-2:50pm</td>
<td>L</td>
<td>Bioinformatic analysis of nucleotide cyclase functional centers and development of ACPred webserver.</td>
<td>Nuo Xu, Changjiang Zhang, Leng Leng Lim and Aloysius Wong.</td>
</tr>
<tr>
<td>2:50-3:15pm</td>
<td>L</td>
<td>Detecting divergent subpopulations in phenomics data using interesting flares.</td>
<td>Methun Kamruzzaman, Ananth Kalyanaraman and Bala Krishnamoorthy.</td>
</tr>
<tr>
<td>3:15-3:30pm</td>
<td>S</td>
<td>Identification of cell types based on word embedding and nonparametric methods in single-cell RNA sequencing data.</td>
<td>Tianyu Wang and Sheida Nabavi.</td>
</tr>
<tr>
<td>3:15-3:30pm</td>
<td>S</td>
<td>Microbiomarkers Discovery in Inflammatory Bowel Diseases using Network-Based Feature Selection.</td>
<td>Mostafa Abbas, Thanh Le, Halima Bensmail, Vasant Honavar and Yasser El-Manzalawy.</td>
</tr>
</tbody>
</table>
## Poster Setup

5:00pm – 6:00pm  
Poster Session Location: Salon E, F  
Poster Pickup by 9:00pm

## Poster Session/Reception

6:00pm – 8:30pm  
Poster Session Location: Salon E, F  
Reception Location: Capitol Foyer

List of accepted posters are at the end of this brochure
Opening and Welcome Remarks

8:45am – 9:00am
Location: Salon E. F. G
General Chair: Amarda Shehu, George Mason University

Keynote Talk 2

9:00am – 10:00am
Location: Salon E, F, G
Huge Cohorts, Genomics, and Clinical Data to Personalize Medicine
Joshua C. Denny, MD, MS, FACMI - Vanderbilt University Medical Center
Session Chair: Hongfang Liu, Mayo Clinic, USA

Abstract: Precision medicine offers the promise of improved diagnosis and for more effective, patient-specific therapies. Typically, such studies have been pursued using research cohorts. At Vanderbilt, we have linked de-identified electronic health records (EHRs), to a DNA repository, called BioVU, which has nearly 250,000 samples. Through BioVU and a NHGRI-funded network using EHRs for discovery, the Electronic Medical Records and Genomics (eMERGE) network, we have used clinical data of genomic basis of disease and drug response using real-world clinical data. The EHR also enables the inverse experiment – starting with a genotype and discovering all the phenotypes with which it is associated – a phenome-wide association study. By looking for clusters of diseases and symptoms through phenotype risk scores, we find unrecognized genetic variants associated with common disease. The era of huge international cohorts such as the UK Biobank, Million Veteran Program, and the newly started All of Us Research Program will make millions of individuals available with dense molecular and phenotypic data. All of Us launched May 6, 2018 and will engage one million diverse individuals across the US who will contribute data and also receive results back.

Biography: Dr. Joshua Denny is Professor of Biomedical Informatics and Medicine, Director of the Center for Precision Medicine and Vice President of Personalized Medicine at Vanderbilt University Medical Center. He is a Fellow of the American College of Medical Informatics and a member of the National Academy of Medicine. He has substantial experience in the design, development, and implementation of electronic health record (EHR) data mining algorithms and was the primary author of several natural language processing systems to support phenotype extraction algorithms for genomic research projects, including development of the phenome-wide association study (PheWAS) method. He is principal investigator (PI) of nodes in the Electronic Medical Records and...
Genomics (eMERGE) Network, Pharmacogenomics Research Network (PGRN), and the Implementing Genomics into Practice (IGNITE) Network. Dr. Denny is PI of the Data and Research Center of the Precision Medicine Initiative All of Us Research Program (previously called the Precision Medicine Initiative Cohort Program), which will eventually enroll at least 1 million Americans in an effort to understand the genetic, environmental, and behavioral factors that influence human health and disease. To date, he has led >100 genome-wide and candidate gene association studies using EHR data linked to genetic data. Dr. Denny serves on a number of mentoring committees and has trained >30 postdoctoral and predoctoral trainees.

10:00am – 10:30am Coffee Break at Capitol Foyer

Paper Session 3

10:30am – Noon

Session 8: Medical Informatics
Location: Salon D
Session Chair: Hagit Shatkay
University of Delaware

1. **H: 10:30-10:55am**

2. **L: 10:25-10:50am**
   Improving Validity of Cause of Death on Death Certificates. Ryan Hoffman, Janani Venugopalan, Li Qu, Hang Wu and May D. Wang.

Session 9: Protein Analysis
Location: Salon E
Session Chair: Sheida Nabavi
University of Connecticut

1. **H: 10:30-10:55am**
   Choosing non-redundant representative subsets of protein sequence data sets using submodular optimization. Libbrecht MW, Bilmes JA, Noble WS.

2. **L: 10:25-10:50am**
   Splice-Aware Multiple Sequence Alignment of Protein Isoforms. Alex Nord, Peter Hornbeck, Kaitlin Carey and Travis Wheeler.

Session 10: Deep Learning and Applications
Location: Salon F
Session Chair: Xiaoqian Jiang
UT Health

1. **L: 10:30-10:55am**
   Cost-Sensitive Deep Active Learning for Epileptic Seizure Detection. Xuhui Chen, JinL: Ji, Tianxi Ji and Pan Li.

2. **L: 10:25-10:50am**

Session 11: Genome Rearrangements and Genetic Variation
Location: Salon G
Session Chair: Layla Oesper
Carleton College

1. **H: 10:30-10:55am**
   Implicit Transpositions in DCJ Scenarios. Avdeyev P, Jiang S, Alekseyev MA.

2. **L: 10:25-10:50am**
   Detecting Chromosomal Inversions from Dense SNPs by Combining PCA and Association Tests. Ronald Nowling and Scott Emrich.
3. **L: 10:50-11:15am**

*CausalTriad: Toward Pseudo Causal Relation Discovery and Hypotheses Generation from Medical Text Data.*

Sendong Zhao, Meng Jiang, Ming Liu, Bing Qin and Ting Liu.

4. **S: 11:15-11:40am**


3. **L: 10:50-11:15am**

*Ensemble Voting Schemes that Improve Machine Learning Models for Predicting the Effects of Protein Mutations.* Sarah Gunderson and Filip Jagodzinski.

3. **L: 10:50-11:15am**


3. **L: 10:50-11:15am**

*A Distributed Disease Subtyping using Somatic Variant Data.* Suzan Arslanturk and Sorin Draghici.

4. **S: 11:15-11:40am**

*Convolutional Neural Networks for Predicting Molecular Binding Affinity to HIV-1 Proteins.* Paul Morris, Yahchayil Dasilva, Evan Clark, William Edward Hahn and Elan Barenholtz.

4. **S: 11:15-11:40am**

*DeepAnnotator: Genome Annotation with Deep Learning.* Mohammad Ruhul Amin, Alisa Yurovsky, Yingtao Tian and Steven Skiena.

---

**Noon – 1:30pm**

Grab Boxed Lunch at Penn Avenue Terrace

---

**Women in Bioinformatics (WIB)**

**12:30pm – 1:30pm**

**Location: Salon F, G**

Session Chair: May D. Wang, Georgia Institute of Technology and Emory University, USA

---

**IEEE/ACM TCBB Editorial Board Meeting**

**2:00pm – 3:30pm**

**Location: State Room**

Session Chair: Aidong Zhang, State University of New York at Buffalo, USA, Co-Chair
Paper Session 4

2:00pm – 3:30pm

Session 12: Medical Informatics
Session Chair: Zhiyong Lu
NIH/NCBI
Location: Salon D

1. H: 2:00-2:25pm

2. L: 2:25-2:50pm
Leveraging Treatment Patterns to Predict Survival of Patients with Advanced Non-Small-Cell Lung Cancer. Kyle Haas, Malika Mahoui, Simone Gupta and Stuart Morto.

3. L: 2:50-3:15pm
Open Information Extraction with Meta-pattern Discovery in Biomedical Literature. Xuan Wang, Yu Zhang, Qi Li, Yinyin Chen and Jiawei Han.

Session 13: Protein Structure and Molecular Dynamics
Session Chair: Christina Boucher
University of Florida
Location: Salon E

1. H: 2:00-2:25pm

2. L: 2:25-2:50pm
clustQ: Efficient protein decoy clustering using superposition-free weighted internal distance comparisons. Rahul Alapati and Debswapna Bhattacharya.

3. L: 2:50-3:15pm

Session 14: Biomarker and Cancer
Session Chair: Tamer Kahveci
University of Florida
Location: Salon F

1. L: 2:00-2:25pm

2. L: 2:25-2:50pm
Integration of Cancer Data through Multiple Mixed Graphical Model. Christopher Ma, Tina Gui, Xin Dang, Yixin Chen and Dawn Wilkins.

3. L: 2:50-3:15pm

Session 15: Phylogenetics
Session Chair: Scott Emrich
University of Tennessee
Location: Salon G

1. L: 2:00-2:25pm
Phylogenetic Consensus for Exact Median Trees. Pawel Tabaszewski, Pawel Gorecki and Oliver Eulenstein.

2. L: 2:25-2:50pm

3. L: 2:50-3:15pm
An Integer Linear Programming Solution for the Domain-Gene-Species Reconciliation Problem. Lei Li and Mukul S. Bansal.
4. S: 3:15-3:30pm
What Happens When?
Interpreting Schedule of Activity Tables in Clinical Trial Documents. Murtaza Dhuliawala, Nicholas Fay, Daniel Gruen and Amar Das.

4. S: 3:15-3:30pm
Hybrid Spectral/Subspace Clustering of Molecular Dynamics Simulations. Ivan Syzonenko and Joshua Phillips.

4. S: 3:15-3:30pm

4. S: 3:15-3:30pm

3:30pm – 4:00pm Coffee Break at Capitol Foyer

NSF-sponsored Student Research Forum

4:00pm – 6:00pm
Location: Salon D
Session Chair: May D. Wang, Georgia Institute of Technology and Emory University, USA

Dinner Banquet

6:00pm – 8:00pm
Location: Salon E, F, G
Saturday, September 1, 2018

8:00am – 8:30am Light Breakfast at Capitol Foyer

Opening and Welcome Remarks

8:45am – 9:00am
Location: Salon E, F, G
General Chair: Cathy Wu, University of Delaware

Keynote Talk 3

9:00am – 10:00am
Location: Salon E, F, G
The Art of Connectivity Mapping
Avi Ma’ayan, PhD - Icahn School of Medicine at Mount Sinai
Session Chair: Jing Li, Case Western Reserve University, USA

Abstract: Motivation: The powerful idea of the Connectivity Mapping proposes the creation of a library of drug induced gene expression signatures. Such a resource can facilitate finding small molecules to mimic or reverse disease signatures, identifying drug targets, discovering the mechanisms of action for novel small molecules, elucidating off-target effect mechanisms, and directing cellular differentiation and reprogramming. A related concept is Gene Set Enrichment Analysis.

Problem statement: In my presentation I will discuss how these two transformative ideas can be expanded in various creative ways to unify knowledge representation in system biology.

Approach: I will demonstrate how expanded Connectivity Mapping and Gene Set Enrichment Analyses combined with Machine Learning can enable imputing and illuminating new biological and pharmacological knowledge.

Biography: The Ma’ayan laboratory applies computational methods to study the complexity of regulatory networks in mammalian cells. We develop software tools to study how molecular intracellular regulatory networks control cellular processes such as differentiation, proliferation, and apoptosis. Our main focus is in developing methods that link changes in genome-wide gene and protein expression to transcriptional regulators and cell signaling pathways. So far we published several popular web-based software tools and databases that enable biologists to perform enrichment analyses and build networks using their own data. By employing the tools we have developed, we work closely with experimental biologists on projects that utilize high-throughput experiments to understand cell regulation at a global scale. So far we published over 130 peer-reviewed articles where several of them are in top-tier journals and are highly cited. In addition, we have mentored over 80 trainees, including postdoctoral fellows, graduate and medical students, undergraduates and high school students. We also contribute to educational and outreach activities of these initiatives including teaching two MOOCs on the Coursera platform. In the past four years we have served as an integral part of the BD2K-LINCS Data Coordination and Integration Center (DCIC), and the Knowledge Management Center (KMC) for the Illuminating
the Druggable Genome (KMC-IDG) project, where we are centrally involved in the development of computational methods, databases and web-based software tools for LINCS, IDG and BD2K. Recently, we also became involved with the construction and design of the NIH Data Commons pilot project.

| 10:00am – 10:30am Coffee Break at Capitol Foyer |

<table>
<thead>
<tr>
<th>Paper Session 5</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Session 16:</strong> Drug Discovery and Docking</td>
</tr>
<tr>
<td><strong>Location:</strong> Salon G</td>
</tr>
</tbody>
</table>
| **Session Chair:** Yuji Zhang   
University of Maryland   
School of Medicine |
| **1. H: 10:30-10:55am** |
| **2. L: 10:55-11:20am** |
| **3. L: 11:20-11:45am** |
| Identifying Experimental Evidence from Biomedical Abstracts Relevant to Drug-Drug Interactions. Gongbo Zhang, Debarati Roychowdhury, Pengyuan Li, Heng-Yi Wu, Shijun Zhang, Lang Li and Hagit Shatkay. |

| **Session 17:** Algorithms |
| **Location:** Salon E |
| **Session Chair:** Byung-Jun Yoon   
Texas A&M University |
| **1. H: 10:30-10:55am** |
| **2. L: 10:55-11:20am** |
| **3. L: 11:20-11:45am** |
| Are Profile Hidden Markov Models Identifiable? Srilakshmi Pattabiraman and Tandy Warnow. |

| **Session 18:** mHealth |
| **Location:** Salon F |
| **Session Chair:** GQ Zhang, Institute of Biomedical Informatics University of Kentucky |
| **1. L: 10:30-10:55am** |
| **2. L: 10:55-11:20am** |
| **3. L: 11:20-11:45am** |
4. S: 11:45am-Noon
Feature Selection to Predict Compound’s Effect on Aging. Hafez Eslami Manoochehri, Susmitha Sri Kadiyala and Mehrdad Nourani.

4. S: 11:45am-Noon
Applying Stochastic Process Model to Imputation of Censored Litudinal Data. Ilya Zhbannikov, Liang He, Konstantin Arbeev and Anatoliy Yashin.

4. S: 11:45am-Noon
Robust Actor-Critic Contextual Bandit for Mobile Health (mHealth) Interventions. Feiyun Zhu, Jun Guo, Ruoyu Li and Junzhou Huang.

12:15pm – 1:45pm Lunch Buffet at Capitol Foyer

Closing Remarks

1:45pm – 2:00pm
Location: Salon F
1. Identification of Non-invasive Cytokine Biomarkers for Polycystic Ovary Syndrome Using Supervised Machine Learning: Daniela Perry, Jeremy Gunawardena, Nicolas Orsi


3. TCGA Lung Cancer Analysis Pipeline: Talip Zengin, Tugba Onal-Suzek

4. Classification of Carbonaceous Nanomaterials based on Patterns of Inflammatory Markers in BAL Fluid and Pathological Outcomes in Lungs: Ishika Desai, William Miller, Vamsi Kodali, Girija Syamlal, Jenny Roberts, Aaron Erdely, Naveena Yanamaala

5. Relation Extraction for Protein-protein Interactions Affected by Mutations: Ziling Fan, Luca Soldaini, Arman Cohan, Nazli Goharian


9. Differences in Engagement Among BabyCenter.com Community Forum Contributors: A Pilot Study: Austin Gu, Casey Taylor

10. A Manifold Learning Based Approach to Reveal the Functional Linkages across Multiple Gene Networks: Nam Nguyen, Ian Blaby, Daifeng Wang

11. Understand Compound Mechanism of Action at a System Biology Scale through Chemical Biology Data Management and Analysis: Huijun Wang, An Chi, Xudong Qiao

12. Profiling Diverse Chemical Space to Map the Druggable Proteome: Huijun Wang, Francisco Garcia, An Chi, Ivan Taracido, Anne Wassermann, Andy Liaw

13. Services4SNPs: A RESTful Platform for Association Rule Mining and Survival Analysis of Genotyping Data: Giuseppe Agapito, Mario Cannataro


15. ProSetComp: A Platform for Protein Set Comparisons: Erdem Turk, Turkan Arit, Delikanli Susus, Ilayda Ucar, Baris Suzek

16. Identifying Symptom Clusters in Women Experiencing Preterm Birth: Ting He, Casey Taylor


18. Practical Feature Selection for Lung Cancer Gene Detection: Min-Wei Hsieh, Hayato Ohwada, Sheng-I Chen


20. PTM Knowledge Networks and LINCS Multi-Omics Data for Kinase Inhibitor Drug-Analitics in Lung Cancer: Xu Zhang, Karen Ross, Tapan Maity, Jake Jaffe, Cathy Wu, Udayan Guha


22. Identifying Genes to Predict Cancer Radiotherapy-Related Fatigue with Machine-Learning Methods: Wei Du, Kristin Dickinson, Calvin Johnson, Leorey Saligan

23. HarMinMax: Gabriel Wright, Anabel Rodriguez, Patricia Clark, Scott Emrich

24. Use of the Informatics for Integrating Biology and the Bedside (i2b2) Population to Test Serum Bilirubin Levels and Risk for Inflammatory Bowl Diseases and the Involvement of Uridine Glucuronosyltransferase Genes:

25. Using a Network Approach to Identify Genes Important during Anther Development: Saleh Tamim, Blake Meyers


28. Early Prediction of Risk Adjustment Factor (RAF) for Medicare Advantage (MA) and Evaluation of its Benefits: Hadi Zarkoob, Prasanna Desikan, Prakash Menon, Harshna Kapashi, Hossein Fakhrai-Rad, Barrie Bradley

29. Comparative Analysis of Tools for Predicting the Functional Impact of mtDNA Variants: Madeline Griffin, Catherine Welsh


31. Faster Computation of Genome Mappability: Sahar Hooshmand, Paniz Abedin, Daniel Gibney, Srinivas Aluru, Sharma Thakachan

32. A Novel Approach for Increasing Taxonomic Resolution in Protein-Based Alignments: Cooper Park, Keir Macartney, Junfu Shen, Kunpeng Xie, Xin Zhang, R. Daniel Bergeron, W. Kelley Thomas, Cheryl Andam, Anthony Westbrook

33. Using Similarity Metrics on Real World Data to Recommend the Next Treatment: Kyle Haas, Malika Mahoui, Simone Gupta, Stuart Morton

34. An Alignment-free Heuristic for Fast Sequence Comparisons with Applications to Phylogeny Reconstruction: Jodh Pannu, Sriram Chockalingam, Sharma Thakachan, Srinivas Aluru

35. RESTful API for iPTMnet: An Integrated Resource for Protein Post-translational Modification Network Discovery: Sachin Gavali, Julie Cowart, Chuming Chen, Karen Ross, Cathy Wu

36. Systematically Prioritizing Targets in Genome-Based Drug Repurposing: Anup Challa, Robert Lavieri, Judith Lewis, Nicole Zaleski, Jana Shirey-Rice, Paul Harris, Jill Pulley

37. Genet-CNV: Boolean Implication Networks for Modelling Genome-Wide Co-occurrence of DNA Copy Number Variations: Salvi Singh, Nancy Guo

38. Coevolutionary Patterns in HIV-1 Pre-integration Complex Offer Insights into Protein Structural Constraints: Madara Hetti Arachchilage, Helen Piontkivska

39. Efficient Distance Calculations Between Genomes Using Mathematical Approximation: Yana Hrytsenko, Noah Daniels, Rachel Schwartz

40. RNA-Seq Dose Response Experiments for Quantification of Off-Target Effects with RNAi Therapeutics: Joseph Barry, Svetlana Morskaya, Tuyen Nguyen, Sarah Solomon, Kevin Fitzgerald, Stuart Milstein, Greg Hinkle


42. Diagnosing Schizophrenia: A Deep Learning Approach: Justin Barry, Srivathsan Srinivasagopalan, Sharma Thakachan, Varadraj Gurupur

43. A Rapid Exact Solution for the Guided Genome Halving Problem: Anton Nekhai, Maria Atamanova, Pavel Avdeyev, Max Alekseyev

44. State Diagrams for Automating Disease “Risk Pyramid” Data Collection and Tailored Clinical Decision Support: Duwayne Willett, Ambarish Pandey, NNeke Ifejika, Vaishnavi Kannan, Jarett Berry, Mujeeb Basit
## Co-Chairs

**Aidong Zhang, Co-Chair**  
State University of New York at Buffalo, USA  

**May D. Wang, Co-Chair**  
Georgia Institute of Technology & Emory University, USA

## Committee Members

<table>
<thead>
<tr>
<th>Name</th>
<th>Institution</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>Srinivas Aluru</td>
<td>Georgia Institute of Technology, USA</td>
<td></td>
</tr>
<tr>
<td>Christopher C. Yang</td>
<td>Drexel University, USA</td>
<td>USA</td>
</tr>
<tr>
<td>Tamer Kahveci</td>
<td>University of Florida, USA</td>
<td></td>
</tr>
</tbody>
</table>

## Program Committee Members

<table>
<thead>
<tr>
<th>Name</th>
<th>Institution</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mukul S. Bansal</td>
<td>University of Connecticut, USA</td>
<td>USA</td>
</tr>
<tr>
<td>Serdar Bozdag</td>
<td>Marquette University</td>
<td>USA</td>
</tr>
<tr>
<td>Broňa Brejová</td>
<td>Comenius University in Bratislava, Slovakia</td>
<td></td>
</tr>
<tr>
<td>Renzhi Cao</td>
<td>Pacific Lutheran University, USA</td>
<td>USA</td>
</tr>
<tr>
<td>Cedric Chauve</td>
<td>Simon Fraser University, Canada</td>
<td>Canada</td>
</tr>
<tr>
<td>Huiyuan Chen</td>
<td>Case Western Reserve University, USA</td>
<td>USA</td>
</tr>
<tr>
<td>Wen Cheng</td>
<td>North Dakota State University, USA</td>
<td>USA</td>
</tr>
<tr>
<td>Rayan Chikhi</td>
<td>CNRS, France</td>
<td>France</td>
</tr>
<tr>
<td>Leonid Chindelevitch</td>
<td>Simon Fraser University, Canada</td>
<td>Canada</td>
</tr>
<tr>
<td>Francesca Cordero</td>
<td>University of Torino, Italy</td>
<td>Italy</td>
</tr>
<tr>
<td>Mohammed El-Kebir</td>
<td>University of Illinois at Urbana-Champaign, USA</td>
<td>USA</td>
</tr>
<tr>
<td>Terry Gaasterland</td>
<td>University of California San Diego, USA</td>
<td>USA</td>
</tr>
<tr>
<td>Dario Ghersi</td>
<td>University of Nebraska at Omaha, USA</td>
<td>USA</td>
</tr>
<tr>
<td>Iman Hajirasouliha</td>
<td>Cornell University, USA</td>
<td>USA</td>
</tr>
<tr>
<td>Steven Hart</td>
<td>Mayo Clinic, USA</td>
<td>USA</td>
</tr>
<tr>
<td>Matthew Hayes</td>
<td>Xavier University of Louisiana, USA</td>
<td>USA</td>
</tr>
<tr>
<td>Lucian Ilie</td>
<td>University of Western Ontario, Canada</td>
<td>USA</td>
</tr>
<tr>
<td>Sorin Istrail</td>
<td>Brown University, USA</td>
<td>USA</td>
</tr>
<tr>
<td>John Kececiogu</td>
<td>University of Arizona, USA</td>
<td>USA</td>
</tr>
<tr>
<td>Paul Kingsbury</td>
<td>Mayo Clinic, USA</td>
<td>USA</td>
</tr>
<tr>
<td>Dan Knights</td>
<td>University of Minnesota, USA</td>
<td>USA</td>
</tr>
<tr>
<td>Mehmet Koyuturk</td>
<td>Case Western Reserve University, USA</td>
<td>USA</td>
</tr>
<tr>
<td>Benjamin Langmead</td>
<td>Johns Hopkins University, USA</td>
<td>USA</td>
</tr>
<tr>
<td>Name</td>
<td>Institution</td>
<td>Location</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>--------------------------------------------------</td>
<td>-------------------</td>
</tr>
<tr>
<td>Ryan Layer</td>
<td>University of Utah, USA</td>
<td>USA</td>
</tr>
<tr>
<td>Heewook Lee</td>
<td>Carnegie Mellon University, USA</td>
<td></td>
</tr>
<tr>
<td>Pan Li</td>
<td>Case Western Reserve University, USA</td>
<td>USA</td>
</tr>
<tr>
<td>Yu Lin</td>
<td>Australian National University, Australia</td>
<td></td>
</tr>
<tr>
<td>Kevin Liu</td>
<td>Michigan State University, USA</td>
<td>USA</td>
</tr>
<tr>
<td>Xiaowen Liu</td>
<td>Indiana University-Purdue University Indianapolis, USA</td>
<td>USA</td>
</tr>
<tr>
<td>Yongchao Liu</td>
<td>Ant Financial, China</td>
<td></td>
</tr>
<tr>
<td>Stefano Lonardi</td>
<td>University of California Riverside, USA</td>
<td></td>
</tr>
<tr>
<td>Jake Luo</td>
<td>University of Wisconsin Milwaukee, USA</td>
<td>USA</td>
</tr>
<tr>
<td>Wenxiu Ma</td>
<td>University of California Riverside, USA</td>
<td>USA</td>
</tr>
<tr>
<td>Ion Mandoiu</td>
<td>University of Connecticut, USA</td>
<td>USA</td>
</tr>
<tr>
<td>Tobias Marschall</td>
<td>Saarland University</td>
<td>USA</td>
</tr>
<tr>
<td>Susan Mcroy</td>
<td>University of Wisconsin Milwaukee, USA</td>
<td>USA</td>
</tr>
<tr>
<td>Tijana Milenkovic</td>
<td>University of Notre Dame, USA</td>
<td>USA</td>
</tr>
<tr>
<td>Franco Milicchio</td>
<td>Roma Tre University, Italy</td>
<td></td>
</tr>
<tr>
<td>Laurent Mouchard</td>
<td>University of Rouen, France</td>
<td></td>
</tr>
<tr>
<td>Robert Patro</td>
<td>Stony Brook University, USA</td>
<td></td>
</tr>
<tr>
<td>Mattia Prosperi</td>
<td>University of Florida, USA</td>
<td></td>
</tr>
<tr>
<td>Sanguthevar Rajasekaran</td>
<td>University of Connecticut, USA</td>
<td></td>
</tr>
<tr>
<td>Stephen Ramsey</td>
<td>Oregon State University, USA</td>
<td></td>
</tr>
<tr>
<td>Anna Ritz</td>
<td>Reed College, USA</td>
<td></td>
</tr>
<tr>
<td>Sergio Salinas</td>
<td>Wichita State University, USA</td>
<td></td>
</tr>
<tr>
<td>Leena Salmela</td>
<td>University of Helsinki, Finland</td>
<td></td>
</tr>
<tr>
<td>Russell Schwartz</td>
<td>Carnegie Mellon University, USA</td>
<td></td>
</tr>
<tr>
<td>Vishakha Sharma</td>
<td>Georgetown University, USA</td>
<td></td>
</tr>
<tr>
<td>Feichen Shen</td>
<td>Mayo Clinic, USA</td>
<td></td>
</tr>
<tr>
<td>Xinghua Shi</td>
<td>University of North Carolina Charlotte, USA</td>
<td></td>
</tr>
<tr>
<td>Ozlem Uzuner</td>
<td>George Mason University, USA</td>
<td></td>
</tr>
<tr>
<td>Henry van Den Bedem</td>
<td>Stanford University, USA</td>
<td></td>
</tr>
<tr>
<td>Tomas Vinar</td>
<td>University in Bratislava, Slovakia</td>
<td></td>
</tr>
<tr>
<td>Kavishwar Wagholikar</td>
<td>Harvard University, USA</td>
<td></td>
</tr>
<tr>
<td>Jerome Waldispulh</td>
<td>McGill University, Canada</td>
<td></td>
</tr>
<tr>
<td>Yanshan Wang</td>
<td>Mayo Clinic, USA</td>
<td></td>
</tr>
<tr>
<td>Chen Wang</td>
<td>Mayo Clinic, USA</td>
<td></td>
</tr>
<tr>
<td>Adam Wright</td>
<td>Harvard University, USA</td>
<td></td>
</tr>
<tr>
<td>Yonghui Wu</td>
<td>University of Florida, USA</td>
<td></td>
</tr>
<tr>
<td>Yufeng Wu</td>
<td>University of Connecticut, USA</td>
<td></td>
</tr>
<tr>
<td>Jie Zheng</td>
<td>Nanyang Technological University, Singapore</td>
<td></td>
</tr>
<tr>
<td>Yaoyun Zhang</td>
<td>University of Texas Health Science Center at Houston, USA</td>
<td></td>
</tr>
<tr>
<td>Yuji Zhang</td>
<td>University of Maryland School of Medicine, USA</td>
<td></td>
</tr>
<tr>
<td>Jinbo Xu</td>
<td>Toyota Technological Institute at Chicago, USA</td>
<td></td>
</tr>
<tr>
<td>Lixia Yao</td>
<td>Mayo Clinic, USA</td>
<td></td>
</tr>
<tr>
<td>Wei Zhang</td>
<td>University of Central Florida, USA</td>
<td></td>
</tr>
<tr>
<td>Yaoyun Zhang</td>
<td>University of Houston, USA</td>
<td></td>
</tr>
<tr>
<td>Jie Zheng</td>
<td>Nanyang Technological University, Singapore</td>
<td></td>
</tr>
</tbody>
</table>
Message from the General Chairs

It is our great pleasure to welcome you to the 9th ACM Conference on Bioinformatics, Computational Biology, and Health Informatics (ACM-BCB 2018). ACM-BCB is the flagship conference of SIGBio, the ACM Special Interest Group in Bioinformatics, Computational Biology, and Biomedical Informatics. Continuing the annual tradition, the conference focuses on interdisciplinary research linking computer science, mathematics, statistics, biology, bioinformatics, biomedical informatics, and health informatics.

With the help of an army of volunteers and the numerous contributions of many researchers like yourself we have put together a program that we hope participants will find very interesting and stimulating. The technical program is the centerpiece of our meeting, with 46 peer-reviewed full research papers and 14 short research papers carefully selected from the 148 original submissions to the conference by diligent efforts of the program committee members. Additionally, we have 60 poster presentations and 11 highlight presentations on recently published, cutting-edge and high-impact research. The technical program is enriched by 7 pre-conference workshops, 1 satellite meeting, and 7 tutorials. The program features diverse sessions, such as the Funding Agency panel, the Hands-on NIH Grant Writing workshop, the NSF Travel Fellowship Awardee Forum, and the Women in Bioinformatics panel.

Such an organization cannot be accomplished without the selfless devotion of numerous volunteers who have tirelessly worked to shape this year’s meeting. We are especially grateful to our program chairs, Christina Boucher, Jing Li, Hongfang Liu, and Mihai Pop, as well as the army of program committee members who volunteered their time to rigorously review submissions. We thank Lenore Cowen and Xiaqian Jiang for selecting high-impact highlight presentations. Special thanks go to Mario Cannataro and Ka Yee Yeung for putting together a strong and diverse workshop program, Rezarta Islamaj Dogan and Casey Overby Taylor for selecting interesting tutorials with broad appeal, and Sheida Navabi and Karen Ross for bringing an outstanding set of timely research for display at the conference as poster presentations. Preetam Ghosh, yet again, handled registrations smoothly, and our proceedings chairs Yang Shen and Xinghua Mindy Shi diligently and meticulously worked with Sheridan Printing and ACM headquarters to produce the conference proceedings in a timely manner.

Special thanks go to Aidong Zhang, the Editor-in-Chief of IEEE/ACM Transactions on Computational Biology and Bioinformatics (IEEE/ACM TCBB), and Dimitrios I. Fotiadis, the Editor-in-Chief of IEEE Journal of Biomedical and Health Informatics for approving ACM-BCB 2018 to have special issues in their journals. We are thankful to May D. Wang for supporting students and selecting NSF travel grant awardees. We are also grateful to Nikos Kyrpides for serving as the Demo and Exhibits Chair, and Debswapna Bhattacharaya, Ercument Cicek, Oznur Tastan, and Pierangelo Veltri for serving as Publicity Chairs. Special thanks go to May D. Wang for continuing the tradition of organizing the Women in Bioinformatics panel, to Nasrin Akhter and Riza Bautista for spearheading student activities, to Aman Sawhney, Sachin Gavali, and Ümit V. Çatalyürek for providing web support, and to many student volunteers for providing logistical and operational assistance in preparation for and during the conference. We also thank the ACM staff, especially John Otero for site negotiations and budget planning, Irene Frawley for conference operations and logistics, and Adrienne Griscti (from ACM) and Lisa Tolles (from Sheridan) for their help in printing the proceedings.
A special note of thanks to our Steering Committee Co-Chairs Aidong Zhang and May D. Wang for being a constant source of support.

We are truly grateful to the National Science Foundation (NSF) for providing travel support for students and young professionals to attend our conference, to interact with distinguished researchers, present their research, network with peers, and build a community of peers. Finally, we would like to thank ACM and SiGBio who provide this platform for researchers to share their state-of-art work in Bioinformatics and Health Informatics.

It has been a privilege to serve as your general chairs by putting the program in place, and making local and hospitality arrangements. On behalf of the entire team of volunteers and the organizing committee, we wish you four days of learning, engagement, pleasant stay, and good fun!

Amarda Shehu
George Mason University

Cathy Wu
University of Delaware
Two new funding opportunities coming soon.

The Chan Zuckerberg Science Initiative will soon be inviting applications for two new funding opportunities:

- **Imaging Scientists Request for Applications**
  This program will support up to 10 Imaging Scientists at imaging centers in the U.S. for 3-5 years. We’re looking for engineers, physicists, mathematicians, computer scientists, or biologists who have focused on developing technology in microscopy or data analysis fields. **Application opens September 5, 2018**

- **Seed Networks for the Human Cell Atlas Request for Applications**
  This program will support 3-year projects to form groups of at least 3 principal investigators to generate data and tools for the Human Cell Atlas, a global effort to map all cells in the human body. **Application opens September 18, 2018**

To stay updated about this and other funding opportunities from CZI Science, please sign up for our mailing list at [http://bit.ly/CZIScience](http://bit.ly/CZIScience)

@CZIScience
chanzuckerberg.com/science
science@chanzuckerberg.com

Chan Zuckerberg Science Initiative

Supporting the science and technology that will make it possible to cure, prevent, or manage all diseases by the end of this century.