ACM-BCB 2019

The 10th ACM Conference on Bioinformatics, Computational Biology, and Health Informatics

Niagara Falls, NY
September 7-10, 2019

ACM BCB 2019 Program

-

Keynote Lecture: 60 minutes (45 minutes for talk and 15 minutes for Q and A)

Highlight Talks: 25 minutes (20 minutes for talk and 5 minutes for Q and A)

Main Conference Regular Paper: 25 minutes (20 minutes for talk and 5 minutes for Q and A)
Main Conference Short Paper: 15 minutes (12 minutes for talk and 3 minutes for Q and A)

By the Numbers

September 7:
- 7 workshops
- 7 tutorials

September 8 – September 10:
- 3 keynotes
- 5 highlights
- 42 regular papers
- 19 short papers

Registration	07:30-17:00 at Foyer
Continental Breakfast	07:30-08:00 at Foyer
Coffee Break	09:00-11:00 & 14:00-16:00 at Foyer

Saturday, September 7, 2019 – Workshops			
Workshop (W)	<u>08:00-12:00</u>	<u>12:00-13:00</u>	<u>13:00-17:30</u>
W: CNB-MAC *Venue: Olmsted*	CNB-MAC		CNB-MAC
W: *CSBW* *Venue: Whitney*	CBSW		CBSW
W: CAME *Venue: Schoellkopf*	CAME	**Lunch at Castellani**	CAME
W: PARBIO *Venue: Adams*			PARBIO
W:MMM Venue: *Hennepin*	MMM		MMM
W:MODI *Venue: Adams*	MODI		

Workshops

W: The Sixth International Workshop on Computational Network Biology: Modeling, Analysis, and Control (CNB-MAC 2019)

Byung-Jun Yoon and Xiaoning Qian, Texas A&M University, United States;
Tamer Kahveci, University of Florida, United States;
Ranadip Pal, Texas Tech University, United States

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 the 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 organize the Sixth International Workshop on Computational Network Biology: Modeling, Analysis, and Control (CNB-MAC 2019) in conjunction with ACM-BCB 2019. The previous CNB-MAC workshops have been successfully held in conjunction with ACM-BCB 2014, ACM-BCB 2015, ACM-BCB 2016, ACM-BCB 2017, and ACM-BCB 2018, 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 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 and fifth CNB-MAC workshops 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.

W: The 2019 Computational Structural Bioinformatics Workshop (CSBW 2019)

Nurit Haspel, UMass Boston, United States;
Dong Si, University of Washington Bothell, United States;
Lin Chen, Elizabeth City State University, United States

Abstract: The 2019 Computational Structural Bioinformatics Workshop will be held in conjunction with ACM-BCB. The rapid accumulation of macromolecular structures presents a unique set of challenges and opportunities in the analysis, comparison, modeling, and prediction of macromolecular structures and interactions. This workshop aims to bring together researchers with expertise in bioinformatics, computational biology, structural biology, data mining, machine learning, optimization, and high-performance computing to discuss new results, techniques, and research problems in computational structural bioinformatics. Selected submissions will be invited to publish extended versions of their

papers in a special issue in MDPI Molecules. Journals used in previous years included the International Journal of Data Mining and Bioinformatics (2007), BMC Structural Biology (2009, 2012) the Journal of Bioinformatics and Computational Biology (2011), Journal of Computational Biology (2015, 2016) and Molecules (2017, 2018).

W: 8th Workshop on Computational Advances in Molecular Epidemiology (CAME 2019)

Yury Khudyakov, Centers for Disease Control and Prevention, United States;
Ion Mandoiu, University of Connecticut, United States;
Pavel Skums and Alex Zelikovsky, Georgia State University, United States

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 the 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 the application of mathematical and computational approaches to molecular epidemiology.

W: 8th International Workshop on Parallel and Cloud-based Bioinformatics and Biomedicine (ParBio)

Mario Cannataro, University Magna Graecia, Catanzaro, Italy;
Wes J. Lloyd, University of Washington, Tacoma, United States;
Giuseppe Agapito, University Magna Graecia, Catanzaro, Italy

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, there is a main bottleneck that involves the storage, integration, and analysis of experimental data, as well as their correlation and integration with publicly available data banks. While parallel computing and Grid computing may offer the computational power and the storage to face this overwhelming availability of data, Cloud Computing is a key technology to hide the complexity of computing infrastructures, to reduce the cost of the data analysis task, and especially to change the overall model of biomedical research and health provision.

High-performance 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).

The Cloud Computing technology, that is able to offer scalable costs and increased reachability, availability and easiness of application use, and possibility to enforce collaboration among scientists, is already changing the business model in different sectors and now it has begun to be adopted in the bioinformatics 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.

The goal of ParBio 2019 is to bring together scientists in the fields of high performance and cloud computing, computational biology and medicine to discuss, among others, the parallel implementation of bioinformatics and biomedical applications and problems and opportunities of moving biomedical and health applications on the cloud. Moreover, big data analytics issues in healthcare and bioinformatics will be addressed. The workshop will focus on research issues, problems and opportunities of moving biomedical and health applications on the cloud, as well as on the opportunity to define guidelines and minimum requirements for a Biomedical Cloud. Moreover, the workshop will discuss about parallel and distributed management and analysis of molecular and clinical data that more and more need to be integrated and analyzed in a joint way.

W: Workshop on Microbiomics, Metagenomics, and Metabolomics (MMM)

Soha Hassoun, Tufts University, United States;
Yasser El-Manzalawy, Geisinger Health System, Pennsylvania State University, United States;
Georg Gerber, Harvard Medical School, Massachusetts Host-Microbiome Center, and Brigham and Women's Hospital, United States;
David Koslicki, Pennsylvania State University, United States;
Gail Rosen, Drexel University, United States

Microbiota are ecological communities of microorganisms found throughout nature. In humans and animals, microbiota communities can reside on or within the body, and exist in a commensal or mutualistic relationship with their host to impact physiological functions and play critical roles in the host's development. These microbial communities can be very complex. One such example is the intestinal microbiota, comprising hundreds of species that interact with other microorganisms in the community as well as their host. Recent studies have demonstrated that microbiota impacts a wide range of physiological processes, including digestion, development of the immune system, and inflammation. Further, significant alterations in the intestinal microbiota composition has shown to correlate with several diseases, including obesity diabetes, cancer, asthma, and even autism spectrum disorder. Characterizing the microbiota and understanding its relation to health and disease stand to significantly improve human health.

Efforts to characterize microbiota have greatly benefited from technical advances in DNA sequencing. In particular, low-cost culture-independent sequencing has made metagenomic and metatranscriptomic surveys of microbial communities practical, including bacteria, archaea, viruses, and fungi associated with the human body, other hosts, and the environment. The resulting data have stimulated the development of many new computational approaches to meta-omic sequence analysis, including metagenomic assembly, microbial identification, and gene, transcript, and pathway metabolic profiling. Further, recent advances in untargeted metabolomics have stimulated the development of many tools that enhance the functional profiling of microbial communities.

W: Machine Learning Models for Multi-omics Data Integration (MODI)

Abed Alkhateeb, University of Windsor, Canada;
Luis Rueda, University of Windsor, Canada

A peer-reviewed proceedings workshop in cutting-edge machine learning approaches and applications in multi-omics data in which researchers in the field showcase and discuss their advanced approaches. The workshop will be half-day long of oral presentations of the accepted papers. We are aiming to host approximately 9 to 12 high-quality accepted works in the field. Each talk will last approximately 15 to 20 minutes, including question/answer session. A coffee with snack break will take place in the middle of the workshop for refreshment, discussions and networking.

The advancement in genome sequencing has helped reveal relevant information about genomic variants in protein functions, spectrums and diseases. Integrative approaches using machine learning and deep learning are applied to rebuild system biology networks of multi-omics including but not limited to DNA and RNA variants (SNPs, indels, CNA, CNV and exons, among others), protein-protein interactions networks and clinical information. Current techniques focus on integrating different molecules to (1) predict the outcomes of diseases such as survivability, progression, and type/subtype of the disease; (2) understand the behavior of molecules and build protein-protein interactions to create or repurpose drugs, in the context of precision medicine. However, the contribution of those different molecules must be deeply analyzed to target the cause rather than just the correlated factors of those molecules. The underlying computational models are aimed to learn the weights of the relationships and contributions of these different omics.

W: Biological ontologies and knowledge bases (BiOK)

Jin Chen, University of Kentucky, United States;
Jiajie Peng, Northwestern Polytechnical University, China

In "Omics" era of the life sciences, it is cost-effective to collect diverse types of genome wide data, which represent the information at various levels of biological systems, including data about genome, transcriptome, epigenome, proteome, metabolome, molecular imaging, molecular pathways, different population of people and clinical/medical records. Currently, big challenge is to represent and use the knowledge contained in the massive data.

A bio-ontology provides standardized and structured vocabulary terms for the scientific community to describe biomedical entities in a domain. In recent years, numerous biomedical ontologies have been developed to represent knowledge about anatomy, molecular function, human phenotype, disease, clinical diagnosis and other areas. Biomedical Ontologies have been proven very useful for knowledge representation, entity annotation, data sharing and data integration et al. in biomedical research.

Knowledge bases are increasingly being used to extract deep biological knowledge and understanding from massive biological data. Knowledge bases can provide information on underlying mechanisms, which statistical inference methods cannot gain insight into. This improvement is largely due to knowledge bases providing a validated biological context for interpreting the ocean of omics.

The biomedical ontologies and knowledge bases workshop provides a vibrant environment for researchers to share their research findings, report novel methods, and discuss the challenges and opportunities in the related fields.

Saturday, September 7, 2019 –Tutorials						
Tutorial (T)	08:00-10:00	10:00-12:00	12:00-13:00	13:00-15:00	15:00-17:00	17:00-17:20
T: Employing Deep Learning to Study Biomolecules (EDL) *Venue: Governor's*	EDL	EDL	Lunch at Castellani			
T: You wrote it, now get it used: Publishing your software with Galaxy and Bioconda (PGB) *Venue: Governor's*			Lunch at Castellani	PGB	PGB	
T: Low-dimensional Representation of Biological Sequence Data (LRBS) *Venue: Red Jacket*	LRBS		Lunch at Castellani			
T: Extracting structure from contaminated symbolic data (ESSD) *Venue: Red Jacket*		ESSD	Lunch at Castellani			
T: Integer Linear Programming in Computational and Systems Biology (ILP) *Venue: Red Jacket*			Lunch at Castellani	ILP		
T: Causal Inference in Biomedical Data Analytics: Basics and Recent Advances (CIBD) *Venue: Red Jacket*			Lunch at Castellani		CIBD	CIBD
T: Machine learning for biomarker discovery in cancer pharmacogenomics data (MLP) *Venue: Cascade 2*			Lunch at Castellani	MLP	MLP	

Tutorials

T: Employing Deep Learning to Study Biomolecules (EDL)

Daniel Veltri (National Institutes of Health) and Kevin Molloy (James Madison University)

Abstract: Deep learning and neural networks are at the frontier of machine learning and artificial intelligence. The applications of these methods are vast and include many applications within bioscience. Most recently, Google's Deep Mind project stunned the protein structure prediction community with their performance in the last critical assessment of structural prediction (CASP) competition.

The objective of this tutorial is three-fold. First, the tutorial will introduce students and researchers that attend ACM-BCB to the deep learning framework Keras. This library, which is built on top of Google's TensorFlow, will be shown utilizing the Python and R programming languages. Second, the tutorial will allow attendees to learn the basic concepts of convolutional neural networks, recurrent neural networks, and transfer learning. Hands on examples and sample code will be provided for each of these separate topics using classic example problems (such as image recognition/classification and text-mining/ sentiment analysis). Third, a hands-on example of using multiple concepts collectively will be employed through an in class competition to identify peptides exhibiting antimicrobial properties. An accompanying website allows attendees to upload and evaluate their models on the fly and further see how it ranks/compares to others.

T: Low-dimensional Representation of Biological Sequence Data (LRBS)

Richard Tillquist (University of Colorado, Boulder)

Abstract: Systems of interest in bioinformatics and computational biology tend to be large, complex, interdependent, and stochastic. As our ability to collect sequence data at finer resolutions improves, we can better understand and predict system behavior under different conditions. Machine learning algorithms are a powerful set of tools designed to help with this understanding. However, many of the most effective of these algorithms are not immediately amenable to application on symbolic data. It is often necessary to map biological symbols to real vectors before performing analysis or prediction using sequence data. This tutorial will cover several techniques for embedding sequence data. Common methods utilizing k-mer count vectors and binary vector representations will be addressed along with state of the art methods based on neural networks, like BioVec, and graph embeddings, like Node2Vec and multilateration. Slides, datasets, and code from the tutorial will be made freely available for future use on GitHub. The materials for this tutorial have been partially funded by the NSF ISS BIGDATA grant No. 1836914.

T: Extracting structure from contaminated symbolic data (ESSD)

Antony Pearson (University of Colorado, Boulder)

Abstract: Symbolic data is the epitome of modern biological datasets. Modern sequencing technologies produce millions of reads giving insights on genome sequence, transcription levels, epigenetic modifications, and much more. To analyze those sequences one usually makes assumptions on their underlying structure, e.g., that the number of reads is Poisson, or that transcription factor binding events are independent at nonoverlapping promoters. These types of assumptions are often not exactly correct in reality. In fact, even when they are valid, a small amount of data

"contamination"' may make them appear untrue. The traditional approach to questioning assumptions on data has been hypothesis testing. This approach has various shortcomings however, particularly it does not give room for a null hypothesis to be "approximately true".' This tutorial introduces a statistical methodology to assess assumptions on symbolic data that may be contaminated. It will demonstrate the applicability of this rather new methodology with publicly available DNA methylation data from ENCODE to question the common but unconscious assumption that methylation of CpGs is exchangeable. Data and code for this tutorial, in the form an iPython Notebook, will be made available via GitHub.

T: Machine learning for biomarker discovery in cancer pharmacogenomics data (MLP)

Arvind Singh Mer, Petr Smirnov and Benjamin Haibe-Kains (University of Toronto)

Abstract: Over the past decade there has been an explosion in the availability of massive datasets combining drug screening with high-throughput molecular profiling in cancer model systems. These datasets have become a rich community resource which can be leveraged for biomarker discovery, in-silico validation, drug repurposing, drug method of action prediction, and to train statistical machine learning models for drug response prediction. However, this data poses unique challenges during analysis and requires methods that are robust to the noise inherent in the drug sensitivity assays. Furthermore, irreproducibility of some findings across studies strongly motivates integrative analysis across studies. Fortunately, tools have been developed implementing bioinformatics and machine learning methods designed specifically for the analysis of pre-clinical pharmacogenomics data.

In this tutorial, participants will become familiar with common preclinical cancer models (such as cell-line, patient derived xenografts and organoids) and publicly available large pharmacogenomics datasets. Next, in the hands on session, they will be introduced to the tools and packages published for analysis of these datasets, with a focus on tools written in R. Furthermore, after becoming familiar with the challenges posed by the noise in the pharmacological assays observed in high-throughput pharmacogenomics, participants will gain hands on experience using these datasets for the purpose of biomarker discovery and validation as well as building machine learning models predictive of drug response. A focus will be on translational research, validating discoveries from in vitro datasets using in vivo pharmacogenomic and clinical datasets. The hands on sessions will be conducted primarily in R and RStudio.

T: Causal Inference in Biomedical Data Analytics: Basics and Recent Advances (CIBD)

May D. Wang and Hang Wu (Georgia Institute of Technology)

Abstract: Causal questions are being answered every day in the biomedical domain, and have significant impact on biomedical experimentation design, data analysis, and healthcare decision making. It's thus important for biomedical researchers to help answer these questions by applying data-driven causal inference algorithms on large-scale biomedical data.

In this tutorial, we focus on identifying the (quantitative) causal effect of interventions. We will introduce a) introduction of causal inference, and popular frameworks for formulating causal inference; b) basics of causal effect identification algorithms; c) state-of-the-art methods by incorporating deep learning and machine learning; d) applications in biomedical data analytics, as well as challenges and opportunities moving forward.

T: You wrote it, now get it used: Publishing your software with Galaxy and Bioconda (PGB)

Daniel Blankenberg (Cleveland Clinic)

Abstract: You've written software, published the code, and described it in a paper. Now, how do you make your software stand out and actually get used? This tutorial introduces two technologies that can make it easy to deploy by researchers around the world and greatly increase your software's reach.

Bioconda is a platform for packaging and publishing bioinformatics software using Conda. The Conda package manager does what previous language and platform specific packagers (e.g., pip, CPAN, CRAN, Bioconductor, apt-get) have done, but in a language and OS agnostic, and much more streamlined way. Tools in Bioconda are easy for infrastructure providers and other researchers to deploy and use. We will introduce Conda and Bioconda principles, and then guide participants through packaging a tool with Bioconda.

Participants will package their newly created Bioconda tool for Galaxy, a widely deployed platform for data integration and analysis in life science research. We will define and test the Bioconda-encapsulated tool for Galaxy and then publish it in the Galaxy Toolshed, where any Galaxy administrator can then install it with a button click.

This will be hands-on. Please bring a wifi-enabled laptop. Instructors will work with participants to install needed software before the conference.

T: Integer Linear Programming in Computational and Systems Biology (ILP)

Dan Gusfield (University of California, Davis)

Abstract: Integer Linear Programming is a versatile modeling and optimization technique that is increasingly used in computational biology in non-traditional ways, most importantly and inventively as a computational tool and language to model and study biological phenomena, to analyze biological data, and to extract biological insight from the models and the data. Integer linear programming is often very effective in solving *instances* of biological problems on realistic data of current importance, even for hard computational problems that lack a worst-case efficient solution method. The effectiveness of the best modern ILP solvers on problem instances of importance in biology opens huge opportunities and could have a truly transformative effect on computation in biology and perhaps medicine.

The goal of the tutorial is to introduce and detail *modeling* and *solving* of real problems in computational biology using integer linear programming. We will illustrate some concepts using a commercial ILP solver from Gurobi Optimization, to solve specific ILPs that we formulate.

Sunday, September 8, 2019

Registration	07:30-17:00 at Main Registration

08:45-09:00	Opening and Welcome Remarks Session Chairs: Michael Buck
09:00-10:00	**Keynote Speaker:** Mark Borodovsky **Keynote:** "New Machine Learning Algorithms for Genome Annotation" Venue: Cascade 2

10:00-10:30	Coffee Break at Grand Foyer

10:30-12:00	**Session 1**	**Session 2**	**Session 3**	**WABI Session 1**
	Translational Bioinformatics I	**Structural Bioinformatics I**	**Biological Networks**	**Molecules and surfaces**
Session Chair	Leonard McMillan	Nurit Haspel	Anna Ritz	Carl Kingsford
Venue	*Red Jacket*	*Olmsted*	*Governor's*	*Porter DeVeaux*

12:00-13:00	Lunch at Cataract

13:00-14:00	TCBB Board Meeting in Governor's	**WABI Keynote Speaker:** Nadia El-Mabrouk **Keynote:** "Inferring the evolutionary history of gene repertoires" Venue: Cascade 2

14:00-15:30	Session 4	Session 5	Session 6	WABI Session 2
	Translational Bioinformatics II	Structural Bioinformatics II	High-throughput Sequencing Data I	Alignment
Session Chair	Russell Schwartz	Lenore Cowen	Steffen Heber	Brian Chen
Venue	*Red Jacket*	*Olmsted*	*Governor's*	*Porter DeVeaux*

15:30-16:00	Coffee Break at Grand Foyer -- *Setup posters in Cataract*

16:00-17:30	**Join ACM SIGBIO General Meeting at Cascade 2**	**WABI Session 3**
		Read mapping
		Session Chair TBD
		Venue Porter DeVeaux

18:00-22:00	**Poster Session and Reception in Cataract** **(List of accepted posters are at the end of this brochure)**

Keynote Talk

New Machine Learning Algorithms for Genome Annotation

Mark Borodovsky, PhD, Georgia Institute of Technology

Abstract: Rapid accumulation of genomic, transcriptomic and protein information creates new opportunities as well as challenges for integration of OMICS data in genome annotation algorithms. Models of prokaryotic genome organization should account for leaderless transcription, non-Shine-Dalgarno ribosomal binding sites and genes horizontally transferred from other species. Automatic parameterization of these more complex models becomes possible via process of incremental expansion of model architecture & step-wise relaxation of restrictions on subsets of parameters upon moving along the steps of iterative training. Accuracy of prediction of eukaryotic genes with complex exon-intron structures could be improved by integrating process of ab initio gene predictions with search for putative orthologues proteins which footprints on genomic DNA are iteratively used in training and prediction. Due to unevenness of evolutionary conservation along a single amino acid chain, parallel use of spliced alignment algorithms for proteins of the same family allows to identify elements of gene structure encoding conserved domains with higher accuracy. I will talk about new genome annotation algorithms: i/ GeneMarkS-2, a part of PGAP, the prokaryotic genome annotation pipeline developed and implemented at NCBI and ii/ eukaryotic self-training gene finder GeneMark-EP utilizing footprints of orthologous proteins in iterative parameterization of HMM statistical models of genome organization.

Biography: Mark Borodovsky received PhD in Applied Mathematics at the Moscow Institute of Physics and Technology. His thesis project was on developing methods of statistically optimal control for systems with incomplete information. He started research in computational biology at the Moscow Institute of Molecular Genetics in 1985. In 1990, he moved to Georgia Tech to continue original work on the GeneMark family algorithms for structural annotation of prokaryotic and eukaryotic genomes. The long term goal is to infer relationships between structural patterns formed by evolution in linear bimolecular sequence with the 3D biological functions Reaching this goal requires new machine learning algorithms integrating genomic, transcriptomic and protein data. Borodovsky is a Founder of the Bioinformatics graduate programs both at Georgia Tech and, more recently, at the Moscow Institute of Physics and Technology. He served as a Chair of the ACM SIGBio Advisory Board 2010-2015.

WABI Keynote Talk

Inferring the evolutionary history of gene repertoires

Nadia El-Mabrouk, PhD, Université de Montréal

Abstract: During evolution, genes are mutated, duplicated, lost and passed to organisms through speciation or Horizontal Gene Transfer (HGT). In addition, their organization in the genome is modified through inversions, transpositions, translocations and other rearrangement events. Understanding how gene order and content have evolved is essential for deciphering gene functions and interactions, with important biological implications. Ideally, all available information on gene sequence and organization should be considered in a single prediction method. However, gene sequence and gene order information are often considered separately. Indeed, inferring rearrangement events modifying gene organization is the purpose of the genome rearrangement field, while inferring losses, duplication and HGT events modifying gene content is the purpose of the gene tree – species tree reconciliation field. In this presentation, I will discuss this issue and present avenues for developing a unifying approach considering both gene orders and gene trees in the purpose of inferring the evolutionary history of gene repertoires.

Biography: Nadia El-Mabrouk is full professor at the Computer Science Department and member of the Centre de Recherche Mathématiques at the University of Montreal. She holds a Ph.D. in theoretical Computer Science from the University Paris VII, obtained in 1996. Nadia has a longstanding experience in developing algorithms for comparative genomics and especially genome rearrangements, gene tree reconstruction and Gene tree/Species tree reconciliation. She is involved, each year, in the program committee of some of the most popular conferences in computational biology such as RECOMB, ISMB, WABI and APBC. She has organized two RECOMB Comparative Genomics Workshops in Montreal. After chairing the Population Genomics and Molecular Evolution track at ISMB, she is, since 2019, acting as ISMB Proceeding Co-chair. Her research appears in a variety of computer science, bioinformatics and life science journals, among them IEEE/ACM, Molecular Biology and Evolution, Bioinformatics, Nature Scientific Reports and BMC-Genomics.

Monday, September 9, 2019				
Registration	07:30-17:00 at Main Registration			
08:45-09:00	Opening and Welcome Remarks Session Chair: Jian Ma **Keynote Speaker:** Christina Leslie **Keynote:** "Decoding Epigenomic Programs in Immunity and Cancer" Venue: Cascade 2			
09:00-10:00				
10:00-10:30	Coffee Break at Capitol Foyer			
10:30-12:00	**Session 7**	**Session 8**	**Session 9**	**WABI Session 4**
-	**Deep Learning I**	**Medical Informatics I**	**Regulatory Genomics I**	**Genomes I: Sequences**
Session Chair	Debswapna Bhattacharya	Dougu Nam	Mukul Bansal	Brendan Mumey
Venue	*Red Jacket*	*Olmsted*	*Governor's*	*Porter DeVeaux*
12:00-14:00	Grab Boxed Lunch at Cataract **Join Women in Bioinformatics (WiB) at Cascade 2**			
14:00-15:30	**Session 10**	**Session 11**	**Session 12**	**WABI Session 5**
	Deep Learning II	**Medical Informatics II**	**Regulatory Genomics II**	**Genomes II: Rearrangement**
Session Chair	Xiao Luo	Jung Lee	Serdar Bozdag	TBD
Venue	*Red Jacket*	*Olmsted*	*Governor's*	*Porter DeVeaux*

15:30-16:00	Coffee Break at Grand Foyer	
16:00-18:00	**Join Funding Agency Panel at Cascade 2**	**WABI Session 6** **Chromosomes and cells**
Session Chair		Michal Ziv-Ukelson
Venue	*Cascade 2*	*Porter DeVeaux*
18:00-20:00	**Dinner Banquet at Cascade Ballroom**	

Keynote Talk

Decoding Epigenomic Programs in Immunity and Cancer

Christina Leslie, PhD, Memorial Sloan Kettering Cancer Center

Abstract: Dysregulated epigenetic programs are a feature of many cancers, and the diverse differentiation states of immune cells as well as their dysfunctional states in tumors are in part epigenetically encoded. We will present recent analysis work and computational methodologies from our lab to decode epigenetic programs from genome-wide data sets. In a recent collaborative work, we characterized chromatin states governing CD8 T cell dysfunction in cancer and reported that tumor-specific T cells differentiate to dysfunction through two discrete chromatin states: an initial plastic state that can be functionally rescued (i.e. through immunotherapy) and a later fixed state that is resistant to therapeutic reprogramming. We now follow up on this work by presenting a computational methodology to decipher transcriptional programs governing chromatin accessibility and gene expression in normal and dysfunctional T cell responses through a large-scale analysis of published data from mouse tumor and chronic viral infection models. This modeling shows that in all these systems, T cells commit to becoming dysfunctional early after an immune challenge, rather than first mounting and then losing an effector response. Through scRNA-seq analysis, we characterize the phenotypic diversity of this common trajectory from plastic to fixed dysfunction. We will also present a recent collaboration with the Sawyers lab on FOXA1 mutants in prostate cancer, showing that somatic alterations in this pioneer transcription factor lead to altered differentiation programs, through analysis of ATAC-seq and ChIP-seq in mouseprostate organoid systems. Finally, we will describe a novel machine learning approach called BindSpace to leverage massive in vitro TF binding data from SELEX-seq experiments through a joint embedding of DNA k-mers and TF labels, leading to improved prediction of TF binding.

Biography: Christina Leslie did her undergraduate degree in Pure and Applied Mathematics at the University of Waterloo in Canada. She was awarded an NSERC 1967 Science and Engineering Fellowship for graduate study and did a PhD in Mathematics at the University of California, Berkeley, where her thesis work dealt with differential geometry and representation theory. She won an NSERC Postdoctoral Fellowship and did her postdoctoral training in the Mathematics Department at Columbia University in 1999-2000. She then joined the faculty of the Computer Science Department and later the Center for Computational Learning Systems at Columbia University, where she began

to work in computational biology and machine learning. In 2007, she moved her lab to Memorial Sloan Kettering Cancer Center, where she is currently a Member of the Computational and Systems Biology Program as well as a Professor of Physiology, Biophysics, and Systems Biology at Weill Cornell Medical College.

Dr. Leslie is widely known for her work developing computational methods to study the global regulation of gene expression and the dysregulation of gene expression programs in cancer. A major methodological contribution of her lab was the introduction of *k*-mer based string kernels for prediction problems involving biological sequences. In addition, since many layers of gene regulation are mediated by DNA and RNA sequence signals, the Leslie lab has pioneered machine learning strategies to combine sequence and expression data to infer gene regulatory programs.

Funding Agency Panel

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

Veerasamy "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)

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

Aidong Zhang
> Fellow of ACM and IEEE
> Professor of Computer Science and Biomedical Engineering
> University of Virginia
> Former Program Director, National Science Foundation

Tuesday, September 10, 2019				
Registration	07:30-17:00 at Capitol Foyer			
08:45-09:00 09:00-10:00	Opening and Welcome Remarks Session Chair: Xinghua (Mindy) Shi **Keynote Speaker:** Heng Huang **Keynote:** " Large-Scale Machine Learning Algorithms for Biomedical Data Science" Venue: Cascade 2			
10:00-10:30	Coffee Break at Foyer			
10:30-12:00	**Session 13**	**Session 14**	**Session 15**	**WABI Session 7**
	Deep Learning III	**Medical Informatics III**	**High-throughput Sequencing Data II**	**Phylogenomics I: Treelike evolution**
Session Chair	Catie Welsh	Xuan Guo	Dan DeBlasio	TBD
Venue	*Red Jacket*	*Olmsted*	*Governor's*	*Porter DeVeaux*
12:00-13:00	*Lunch at Cataract*			
13:00-14:30				**WABI Session 8** **Phylogenomics II: Non-treelike evolution**
Session Chair				Tandy Warnow
Venue				*Porter DeVeaux*

14:30-15:00	Coffee Break at Foyer		
15:00-16:30	**Session 16**	**Session 17**	**WABI Session 9**
	Bioimages & Function Annotation	**Comparative genomics & cancer phylogenetics**	**Phylogenomics III: Non-treelike evolution**
Session Chair	Pierangelo Veltri	Brendan Mumey	Dan Gusfield
Venue	*Red Jacket*	*Governor's*	*Porter DeVeaux*
16:30-17:00	**Closing Remarks at Cascade**		

Keynote Talk

Large-Scale Machine Learning Algorithms for Biomedical Data Science

Heng Huang, PhD, University of Pittsburgh

Abstract: Data science is accelerating the translation of biological and biomedical data to advance the detection, diagnosis, treatment, and prevention of diseases. However, the unprecedented scale and complexity of large-scale biomedical data have presented critical computational bottlenecks requiring new concepts and enabling tools. To address the challenging problems in current biomedical data science, we proposed several novel large-scale machine learning models for multi-dimensional data integration, heterogeneous multi-task learning, longitudinal feature learning, etc. Meanwhile, to deal with the big data computations, we proposed new asynchronous distributed stochastic gradient and coordinate descent methods for efficiently solving convex and non-convex problems, and also parallelized the deep learning optimization algorithms with layer-wise model parallelism. We applied our new large-scale machine learning models to analyze the multi-modal and longitudinal Electronic Medical Records (EMR) for predicting the heart failure patients' readmission and drug side effects, integrate the neuroimaging and genome-wide array data to recognize the phenotypic and genotypic biomarkers, and detect the histopathological image markers and the multi-dimensional cancer genomic biomarkers in precision medicine studies.

Biography: Dr. Heng Huang is a John A. Jurenko Endowed Professor in Computer Engineering at the Department of Electrical and Computer Engineering at the University of Pittsburgh, and also a Professor in Biomedical Informatics at University of Pittsburgh Medical Center. Dr. Huang received his PhD degree in Computer Science at Dartmouth College. His research areas include machine learning, big data mining, health informatics, medical image analysis, bioinformatics, neuroinformatics, and precision medicine.

BCB Conference Paper Presentations

Session 1	Translational bioinformatics I
Long	Title: ENCORE: a Visualization Tool for Insight into Circadian Omics Authors: Hannah De Los Santos, Kristin P. Bennett and Jennifer M. Hurley
Long	Title: A Visual Analytics Framework for Analysis of Patient Trajectories Authors: Kaniz Fatema Madhobi, Ananth Kalyanaraman, Methun Kamruzzaman, Eric Lofgren, Bala Krishnamoorthy and Rebekah Moehring
Long	Title: Drug Repositioning Predictions by NMTF of Integrated Association Data Authors: Gaëtan Dissez, Gaia Ceddia, Pietro Pinoli, Stefano Ceri and Marco Masseroli
Short	Title: Model-Agnostic Interpretation of Cancer Classification with Multi-Platform Genomic Data Authors: Olatunji Oni and Sanzheng Qiao

Session 2	Structural bioinformatics I
Highlight	Title: Protein Tertiary Structure Modeling Driven by Deep Learning and Contact Distance Prediction in CASP13 Authors: Jianlin Cheng
Long	
	Title: Human Protein Complex Signatures for Drug Repositioning Authors: Fei Wang, Xiujuan Lei, Bo Liao and Fangxiang Wu
Long	
	Title: Majority Vote Cascading: a Semi-Supervised Framework for Improving Protein Function Prediction
Short	Authors: John Lazarsfeld, Jonathan Rodriguez, Mert Erden, Yuelin Liu and Lenore Cowen
	Title: GANDALF: GAN based peptide design Authors: Allison Rossetto and Wenjin Zhou

Session 3	Biological networks
Long	Title: Large-scale Analysis of Drug Combinations by Integrating Multiple Heterogeneous Information Networks Authors: Huiyuan Chen, Sudha K. Iyengar and Jing Li
Long	Title: Robinson-Foulds Reticulation Networks Authors: Alexey Markin, Tavis Anderson, Venkata Sai Krishna Teja Vadali and Oliver Eulenstein
Long	Title: Network-Based Modeling of Sepsis: Quantification and Evaluation of Simultaneity of Organ Dysfunctions

	Authors: Ali Jazayeri, Muge Capan, Christopher Yang, Farzaneh Khoshnevisan, Min Chi and Ryan Arnold
Short	Title: Finding conserved patterns in multilayer networks Authors: Yuanfang Ren, Aisharjya Sarkar, Ahmet Ay, Alin Dobra and Tamer Kahveci

Session 4	Translational bioinformatics II
Long	Title: Multiple Graph Kernel Fusion Prediction of Drug Prescription Authors: Hao-Ren Yao, Der-Chen Chang, Ophir Freider, Wendy Huang and Tian-Shyug Lee
Long	Title: Drug-Drug Interaction Prediction Based on Knowledge Graph Embeddings and Convolutional-LSTM Network Authors: Md. Rezaul Karim, Michael Cochez, Joao Bosco Jares, Mamtaz Uddin, Oya Beyan and Stefan Decker
Short	Title: Feedback Regulation of Immune Response to Maximum Exercise in Gulf War Illness Authors: Cole A. Lyman, Mark Clement, Travis J. A. Craddock, Mary Fletcher, Nancy G. Klimas and Gordon Broderick
Short	Title: Automate the Peripheral Arterial Disease Prediction in Lower Extremity Arterial Doppler Study using Machine Learning and Neural Networks Authors: Lena Ara, Xiao Luo, Alan Sawchuk and David Rollins

Session 5	Structural bioinformatics II
Highlight	Title: How effective is contact-assisted protein threading? Authors: Sutanu Bhattacharya and Debswapna Bhattacharya
Long	Title: Transmembrane Topology Identification by Fusing Evolutionary and Co-evolutionary Information with Cascaded Bidirectional Transformers Authors: Zhen Li, Chongming Ni, Jinbo Xu, Xin Gao, Shuguang Cui and Sheng Wang
Long	Title: Integration of heterogeneous experimental data improves global map of human protein complexes Authors: Jose Lugo-Martinez, Joern Dengjel, Ziv Bar-Joseph and Robert Murphy
Short	Title: Using Sequence-Predicted Contacts to Guide Template-free Protein Structure Prediction Authors: Ahmed Bin Zaman, Prasanna Parthasarathy and Amarda Shehu
Short	Title: Automated Threshold Selection for Cryo-EM Density Maps Authors: Jonas Pfab and Dong Si

Session 6: High-throughput Sequencing Data I	
Long	Title: Practical universal k-mer sets for minimizer schemes Authors: Dan DeBlasio, Fiyinfoluwa Gbosibo, Carl Kingsford and Guillaume Marçais
Long	Title: MitoMut: an efficient approach to detecting mitochondrial DNA deletions from paired-end next-generation sequencing data Authors: C. Shane Elder and Catherine E Welsh
Short	Title: ELITE: Efficiently Locating Insertions of Transposable Elements Authors: Anwica Kashfeen, Harper Fauni, Timothy Bell, Fernando Pardo-Manuel de Villena and Leonard McMillan

September 9, 2019

Session 7 Deep Learning I	
Long	Title: A Deep Learning Approach to Phase-Space Analysis for Seizure Detection Authors: Patrick Luckett, Thomas Watts, J. Todd McDonald, Lee Hively and Ryan Benton
Long	Title: Deep Learning for High-Order Drug-Drug Interaction Prediction Authors: Bo Peng and Xia Ning
Long	Title: Integrative Feature Ranking by Applying Deep Learning on Multi Source Genomic Data Authors: Fariba Khoshghalbvash and Jean Gao
Short	Title: A Sparse Convolutional Predictor with Denoising Autoencoders for Phenotype Prediction Authors: Junjie Chen and Xinghua Shi

Session 8 Medical Informatics I	
Long	Title: PEARL: Prototype Learning via Rule Learning Authors: Tianfan Fu, Tian Gao, Cao Xiao, Tengfei Ma and Jimeng Sun
Long	Title: Influenza-Like Symptom Prediction by Analyzing Self-Reported Health Status and Human Mobility Behaviors Authors: Fenglong Ma, Shiran Zhong, Jing Gao and Ling Bian
Long	Title: Time-series as Background Data for Relating Medical Diagnoses Terms Authors: Saket Gurukar and Srikanta Bedathur
Short	Title: Multi-modal predictive models of diabetes progression Authors: Ramin Ramazi, Christine Perndorfer, Emily Soriano, Jean-Philippe Laurenceau and Rahmatollah Beheshti

Session 9 Regulatory genomics I

Long	Title: Unexpected predictors of antibiotic resistance in housekeeping genes of Staphylococcus aureus Authors: Mattia Prosperi, Taj Azarian, Judith A. Johnson, Marco Salemi, Franco Milicchio and Marco Oliva
Long	Title: Gene Set Databases: A Fountain of Knowledge or a Siren Call? Authors: Farhad Maleki, Katie Ovens, Ian McQuillan, Elham Rezaei, Alan M. Rosenberg and Anthony J. Kusalik
Long	Title: PhenoGeneRanker: A Tool for Gene Prioritization Using Complete Multiplex Heterogeneous Networks Authors: Cagatay Dursun, Naoki Shimoyama, Mary Shimoyama, Michael Schlappi and Serdar Bozdag
Long	Title: Exponentially few RNA structures are designable Authors: Hua-Ting Yao, Cedric Chauve, Mireille Regnier and Yann Ponty

Session 10 Deep Learning II

Long	Title: SAU-Net: A Universal Deep Network for Cell Counting Authors: Yue Guo, Jason Stein, Guorong Wu and Ashok Krishnamurthy
Long	Title: Biomedical Mention Disambiguation using a Deep Learning Approach Authors: Chih-Hsuan Wei, Kyubum Lee, Robert Leaman and Zhiyong Lu
Short	Title: Deep Convolutional Neural Network for Automated Detection of Mind Wandering using EEG Signals Authors: Seyedroohollah Hosseini and Xuan Guo
Short	Title: Extraction of Tumor Site from Cancer Pathology Reports using Deep Filters Authors: Abhishek Dubey, Jacob Hinkle, Blair Christian and Georgia Tourassi

Session 11 Medical Informatics II

Long	Title: NamedKeys: Unsupervised Keyphrase Extraction for Biomedical Documents Authors: Zelalem Gero and Joyce Ho
Long	Title: Learning Electronic Health Records through Hyperbolic Embedding of Medical Ontologies Authors: Qiuhao Lu, Nisansa de Silva, Sabin Kafle, Jiazhen Cao, Dejing Dou, Thien Nguyen, Prithviraj Sen, Brent Hailpern, Berthold Reinwald and Yunyao Li
Long	Title: Network Analysis and Recommendation for Infectious Disease Clinical Trial Research Authors: Magdalyn Elkin, Whitney Andrews and Xingquan Zhu

Session 12 Regulatory Genomics II	
Highlight	Title: Beta-binomial modeling of CRISPR pooled screen data identifies target genes with greater sensitivity and fewer false negatives Authors: Hyun-Hwan Jeong, Seon Young Kim, Maxime Rousseaux, Huda Zoghbi and Zhandong Liu
Long	Title: Predicting G-quadruplexes from DNA sequences using multi-kernel convolutional neural networks Authors: Mira Barshai and Yaron Orenstein
Long	Title: miRDriver: A Tool to Infer Copy Number Derived miRNA-Gene Networks in Cancer Authors: Banabithi Bose and Serdar Bozdag
Short	Title: Pre-Phaser: Precise Cell-Cycle Phase Detector for scRNA-seq Authors: Alisa Yurovsky, Bruce Futcher and Steve Skiena

September 10, 2019

Session 13 Deep Learning III	
Highlight	Title: Low-dimensional representation of genomic sequences Authors: Richard Tillquist and Manuel Lladser
Short	Title: Multivariate Multi-step Deep Learning Time Series Approach in Forecasting Parkinson's Disease Future Severity Progression Authors: Nur Hafieza Ismail, Mengnan Du, Diego Martinez and Zhe He
Short	Title: Dynamic Cluster-based Retrieval and Discovery for Biomedical Literature Authors: Michael Ortiz, Heejun Kim, Mika Wang, Kazuhiro Seki and Javed Mostafa
Short	Title: Extracting Molecular Entities and Their Interactions from Pathway Figures Based on Deep Learning Authors: Fei He, Duolin Wang, Yulia Innokenteva, Olha Kholod, Dmitriy Shin and Dong Xu

Session 14 Medical Informatics III	
Long	Title: Identifying Symptom Clusters in Breast Cancer and Colorectal Cancer Patients using EHR Data Authors: Priyanka Gandhi, Xiao Luo, Susan Storey, Zuoyi Zhang, Zhi Han and Kun Huang
Long	Title: PPPred: Classifying Protein-phenotype Co-mentions Extracted from Biomedical Literature Authors: Morteza Pourreza Shahri, Gillian Reynolds, Mandi M. Roe and Indika Kahanda
Short	Title: Copy Number Variation Detection Using Total Variation Authors: Fatima Zare and Sheida Nabavi
Short	Title: SMILEBERT: Large Scale Unsupervised Pre-Training for Molecular Property Prediction Authors: Sheng Wang, Yuzhi Guo, Yuhong Wang, Hongmao Sun and Junzhou Huang

Session 15 High-throughput Sequencing Data II	
Long	Title: Effective clustering for single cell sequencing cancer data Authors: Simone Ciccolella, Murray Patterson, Paola Bonizzoni and Gianluca Della Vedova
Long	Title: ParRefCom : Parallel Reference-based Compression of Paired-end Genomics Read Datasets Authors: Nagakishore Jammula and Srinivas Aluru
Long	Title: Using a Novel Negative Selection Inspired Anomaly Detection Algorithm to Identify Corrupted Ribo-seq and RNA-seq Samples Authors: Patrick Perkins and Steffen Heber

Session 16 Bioimages & Function Annotation	
Highlight	Title: CoPhosK: A Method for Comprehensive Kinase Substrate Annotation Using Co-phosphorylation Analysis Authors: Marzieh Ayati, Danica Wiredja, Daniela Schlatzer, Sean Maxwell, Ming Li, Mehmet Koyuturk and Mark Chance
Long	Title: Learning to Evaluate Color Similarity for Histopathology Images using Triplet Networks Authors: Anirudh Choudhary, Hang Wu, Li Tong and May Wang
Long	Title: Residual Deep Learning System for Mass Segmentation and Classification in Mammography Authors: Dina Abdelhafiz, Sheida Nabavi, Reda Ammar, Clifford Yang and Jinbo Bi
Long	Title: Fusion in Breast Cancer Histology Classification Authors: Juan Vizcarra, Ryan Place, Tong Li, David Gutman and May Wang

Session 17 Comparative genomics & cancer phylogenetics	
Long	Title: Co-evolving patterns in temporal networks of varying evolution Authors: Rasha Elhesha, Aisharjya Sarkar, Pietro Cinaglia, Christina Boucher and Tamer Kahveci
Long	Title: Scalable Statistical Introgression Mapping Using Approximate Coalescent-Based Inference Authors: Qiqige Wuyun, Nicholas Vankuren, Marcus Kronforst, Sean Mullen and Kevin Liu
Long	Title: On Inferring Additive and Replacing Horizontal Gene Transfers Through Phylogenetic Reconciliation Authors: Misagh Kordi, Soumya Kundu and Mukul S. Bansal
Short	Title: Reconstructing Intra-Tumor Heterogeneity via Convex Optimization and Branch-and-Bound Search Authors: Shorya Consul and Haris Vikalo

WABI Presentations

Sept. 8 (Long means 25min + 5 min for questions, answers, and change over)

WABI Session 1 Molecules and surfaces	
Long	Title: Fast and accurate structure probability estimation for simultaneous alignment and folding of RNAs Authors: Milad Miladi, Martin Raden, Sebastian Will, and Rolf Backofen
Long	
	Title: Quantified uncertainty of the flexible protein-protein docking algorithm
Long	Authors: Nathan Clement
	Title: pClay: A Precise Parallel Algorithm for Comparing Molecular Surfaces Authors: Georgi D. Georgiev, Kevin F. Dodd and Brian Y. Chen

WABI Session 2 Alignment	
Long	Title: Validating Paired-end Read Alignments in Sequence Graphs Authors: Chirag Jain, Haowen Zhang, Alexander Dilthey, and Srinivas Aluru
Long	Title: Bounded-length Smith - Waterman alignment Authors: Alexander Tiskin

WABI Session 3 Read mapping	
Long	Title: Context-Aware Seeds for Read Mapping Authors: Hongyi Xin, Mingfu Shao, and Carl Kingsford
Long	Title: Read Mapping on Genome Variation Graphs Authors: Kavya Vaddadi, Rajgopal Srinivasan, and Naveen Sivadasan

Sept. 9

WABI Session 4 Genomes I: Sequences	
Long	Title: Synteny paths for assembly graphs comparison Authors: Evgeny Polevikov and Mikhail Kolmogorov
Long	Title: Faster pan-genome construction for efficient differentiation of naturally occurring and engineered plasmids with plaster Authors: Qi Wang, R. A. Leo Elworth, Tian Rui Liu, and Todd J. Treangen
Long	Title: Finding all maximal perfect haplotype blocks in linear time Authors: Jarno Alanko, Hideo Bannai, Bastien Cazaux, Pierre Peterlongo, and Jens Stoye

WABI Session 5 Genomes II: Rearrangement	
Long	Title: Detecting Transcriptomic Structural Variants in Heterogeneous Contexts via the Multiple Compatible Arrangements Problem Authors: Yutong Qiu, Cong Ma, Han Xie, and Carl Kingsford
Long	Title: Weighted Minimum-length Rearrangement Scenarios Authors: Pijus Simonaitis, Annie Chateau, and Krister M. Swenson
Long	Title: The future of WABI Presenter: Jens Stoye

WABI Session 6 Chromosomes and cells	
Long	Title: Topological data analysis reveals principles of chromosome structure in cellular differentiation Authors: Natalie Sauerwald, Yihang Shen, and Carl Kingsford
Long	Title: Inferring diploid 3D chromatin structures from Hi-C data Authors: Alexandra Gesine Cauer, Grkan Yardımci, Jean-Philippe Vert, Nelle Varoquaux, and William Stafford Noble
Long	Title: A Combinatorial Approach for Single-cell Variant Detection via Phylogenetic Inference Authors: Mohammadamin Edrisi, Hamim Zafar, and Luay Nakhleh
Long	Title: Jointly embedding multiple single-cell omics measurements Authors: Jie Liu, Yuanhao Huang, Ritambhara Singh, Jean-Philippe Vert, and William Stafford Noble

Sept. 10

WABI Session 7 Phylogenomics I: Treelike evolution	
Long	Title: Building a Small and Informative Phylogenetic Supertree Authors: Jesper Jansson, Konstantinos Mampentzidis, and Sandhya Thekkumpadan Puthiyaveedu
Long	Title: TRACTION: Fast non-parametric improvement of estimated gene trees Authors: Sarah Christensen, Erin Molloy, Pranjal Vachaspati, and Tandy Warnow
Long	Title: Rapidly Computing the Phylogenetic Transfer Index Authors: Jakub Truszkowski, Olivier Gascuel, and Krister M. Swenson

WABI Session 8 Phylogenomics II: Non-treelike evolution	
Long	Title: Empirical Performance of Tree-based Inference of Phylogenetic Networks Authors: Zhen Cao, Jiafan Zhu, and Luay Nakhleh

Long	Title: Consensus Clusters in Robinson-Foulds Reticulation Network Authors: Alexey Markin and Oliver Eulenstein
Long	Title: Better Practical Algorithms for rSPR Distance and Hybridization Number Authors: Kohei Yamada, Zhi-Zhong Chen, and Lusheng Wang

WABI Session 9 Phylogenomics III: Non-treelike evolution	
Long	Title: Alignment- and reference-free phylogenomics with colored de-Bruijn graphs Authors: Roland Wittler
Long	Title: A New Paradigm for Identifying Reconciliation-Scenario Altering Mutations Conferring Environmental Adaptation Authors: Roni Zoller, Meirav Zehavi, and Michal Ziv-Ukelson

BCB and WABI Posters

BCB Posters:

192	Modeling Phytoplankton Movement and Fitness in Lakes	Amy R. Lazarte (Reed); Samuel B. Fey (Reed); Anna Ritz (Reed)
195	Application of Comparative Biosequence Analysis to Understand Antibiotic Resistance in Superbugs	Guinevere Sieradzki (Milwaukee SOE); Sabrina Mierswa (Milwaukee SOE); Jung Lee (Milwaukee SOE)
197	A novel workflow for semi-supervised annotation of cell-type clusters in mass cytometry data	Abhinav Kaushik (Stanford); Diane Dunham (Stanford); Monali Manohar (Stanford); Kari Nadeau (Stanford); Sandra Andorf (Stanford)
198	Evaluation of five sentence similarity models on electronic medical records	Qingyu Chen (NIH); Jingcheng Du (UTHealth); John Wilbur (NIH); Sun Kim (NIH); Zhiyong Lu (NIH)
200	Refinement of G protein-coupled receptor structure models: Improving the prediction of loop conformations and drug binding	Bhumika Arora (IIT Bombay); Venkatesh Kareenhalli (IIT Bombay, Monash U); Patrick Sexton (Monash U)
201	Augmenting Quality Assurance Measures in Radiation Oncology with Machine Learning	Malvika Pillai (UNC); Karthik Adapa (UNC)
202	Toward a sequence-based physicochemical approach to variable-length B-cell epitope prediction for antipeptide paratopes recognizing flexibly disordered targets: insights drawn from protein folding modeled as polymer collapse	Salvador Eugenio Caoili (U of Philippines Manila)
203	Scalable Statistical Introgression Mapping Using Approximate Coalescent-Based Inference	Qiqige Wuyun (MSU); Nicholas Vankuren (U of Chicago); Marcus Kornforst (U of Chicago); Sean Mullen (Boston U); Kevin Liu (MSU)
204	Contact-assisted protein threading: an evolving new direction	Sutanu Bhattacharya (Auburn); Debswapna Bhattacharya (Auburn)
205	Pangenome-Wide Association Studies with Frequented Regions	Buwani Manuweera (Montana State); Indika Kahanda (Montana State); Brendan Mumey (Montana State); Joann Mudge (NCGR); Thiruvarangan Ramaraj (NCGR); Alan Cleary (NCGR)
207	l_0DL: Joint Image Gradient l_0-norm with Dictionary Learning for limited-angle CT	Moran Xu (Southeast U); Dianlin Hu (Southeast U); Weiwen Wu (Chongqing U)
208	Predicting G-quadruplexes from DNA sequences using multi-kernel convolutional neural networks	Mira Barshai (Ben-Gurion); Yaron Orenstein (Ben-Gurion)

209	PubTator Central: Automated Concept Annotation of Biomedical Full Text Articles	Chih-Hsuan Wei (NCBI); Alexis Allot (NCBI); Robert Leaman (NCBI); Zhiyong Lu (NCBI)
210	Majority Vote Cascading: a Semi-Supervised Framework for Improving Protein Function Prediction.	John Lazarsfeld (Tufts); Jonathan Rodriguez (Tufts); Mert Erden (Tufts); Yuelin Liu (Tufts); Lenore Cowen (Tufts)
212	Integration of heterogeneous experimental data improves global map of human protein complexes	Jose Lugo-Martinez (CMU); Jörn Dengjel (Fribourg); Ziv Bar-Joseph (CMU); Robert F. Murphy (CMU)
213	Long Non-coding RNA Based Cancer Classification using Deep Neural Networks	Abdullah Mamun (FIU); Ananda Mondal (FIU)
214	PEARL: Prototype Learning via Rule Learning	Tianfan Fu (GIT); Tian Gao (IBM); Cao Xiao (IQVIA); Tengfei Ma (IBM); Jimeng Sun (GIT)
215	Reducing Redundancy in Biological Sequence Alignment using Cache Optimization	Evan Stene (UC Denver); Farnoush Banaei-Kashani (UC Denver)
216	Simulating Uncertainty of Early Warning Scores in Early Sepsis Detection	Ali Jazayeri (Drexel); Muge Capan (Drexel); Christopher Yang (Drexel); Siddhartha Nambiar (NCSU); Maria Mayorga (NCSU); Julie Ivy (NCSU); Ryan Arnold (Drexel)
218	Pseudotime Based Analysis of Cancer Dynamics	Tasmia Aqila (FIU); Ananda Mondal (FIU)
219	Community Based Cancer Biomarker Identification from Gene Co-expression Network	Raihanul Tanvir (FIU); Mona Maharjan (FIU); Ananda Mondal (FIU)
220	Explanation of Machine Learning Models Using Improved Shapley Additive Explanation	Yasunobu Nohara (Kyushu U); Koutarou Matsumoto (Saiseikai Kumamoto); Hidehisa Soejima (Saiseikai Kumamoto); Naoki Nakashima (Kyusu U)
221	Transmembrane Topology Identification by Fusing Evolutionary and Co-evolutionary Information with Cascaded Bidirectional Transformers	Zhen Li (CUHKSZ); Chongming Ni (Tsinghua U); Sheng Wang (KAUST)
222	Dynamic interaction network inference from longitudinal microbiome data	Jose Lugo-Martinez (CMU); Daniel Ruiz Perez (BioRG); Giri Narasimhan (FIU); Ziv Bar-Joseph (CMU)

WABI Posters:

Getting insight into the pan-genome structure with Pangtree	Paulina Dziadkiewicz, Norbert Dojer
Memory-efficient parallel algorithms for approximating compact universal hitting sets	Baris Ekim

Organizing Committee

General Chairs:
Xinghua (Mindy) Shi, Temple University, USA
Michael Buck, University of Buffalo, USA

Program Chairs:
Jian Ma, Carnegie Mellon University, USA
Pierangelo Veltri, University Magna Graecia of Catanzaro, Italy

Workshop Chairs:
Wei Zhang, University of Central Florida, USA
Xia Ning, The Ohio State University, USA

Tutorial Chairs:
Shaun Mahony, Pennsylvania State University, USA
Simina Boca, Georgetown University, USA

Poster Chairs:
Dong Si, University of Washington at Bothell, USA
Daniel Mcskimming, University of Buffalo, USA

Demo and Exhibit Chair:
Robert (Bob) W. Cottingham, Oak Ridge National Laboratory

Highlights Chairs:
Jianlin Cheng, University of Missouri, USA

Registration Chair:
Preetam Ghosh, Virginia Commonwealth University, USA

Student Travel Award Chair:
May D. Wang, Georgia Institute of Technology and Emory University, USA

Women in Bioinformatics (WiB) Chair:
May D. Wang, Georgia Institute of Technology and Emory University, USA

Proceedings Chairs:
Yang Shen, Texas A&M University, USA

Publicity Chairs:
Pietro Hiram Guzzi, University of Catanzaro, Italy
Paolo Soda, University Campus Bio-Medico di Roma, Italy
Mattia Prosperi, University of Florida, USA

Web Admin:
Letu Qingge, University of North Carolina at Charlotte, USA
Julien Kann, University of Buffalo, USA

Steering Committee:
Aidong Zhang, University of Virginia, USA, Co-Chair
May D. Wang, Georgia Institute of Technology and Emory University, USA, Co-Chair
Srinivas Aluru, Georgia Institute of Technology, USA
Christopher C. Yang, Drexel University, USA
Tamer Kahveci, University of Florida, USA

Program Committee Members

Giuseppe Agapito, *Università Magna Graecia Catanzaro*
Can Alkan, *Bilkent University*
Ferhat Ay, *La Jolla Institute for Immunology*
Mukul S. Bansal, *University of Connecticut*
Nematollah K. Batmanghelich, *University of Pittsburgh*
Simina Boca, *Georgetown University*
Cedric Chauve, *Simon Fraser University*
Jianlin Cheng, *University of Missouri Columbia*
Rayan Chikhi, *Pasteur Institute*
Leonid Chindelevitch, *Simon Fraser University*
Luis Pedro Coelho, *Fudan University*
Francesca Cordero, *University of Torino*
Juan Cortés, *LAAS-CNRS*
Marie-Dominique Devignes, *LORIA-CNRS*
Mohammed El-Kebir, *University of Illinois at Urbana-Champaign*
Amin Emad, *McGill University*
Petko Fiziev, I*llumina AI Lab*
Preetam Ghosh, *Virginia Commonwealth University*
Raluca Gordan, *Duke University*
Jun-Tao Guo, *University of North Carolina - Charlotte*
Pietro Hiram Guzzi, *University Magna Graecia of Catanzaro*
Faraz Hach, *University of British Columbia and Vancouver Prostate Centre*
Iman Hajirasouliha*, Weill Cornell Medicine - Cornell University*
Fereydoun Hormozdiari, *University of Washington*
Heng Huang, *University of Pittsburgh*
Tamer Kahveci*, University of Florida*
John Kececiogu, *University of Arizona*
Sun Kim, *Seoul National University*
Smita Krishnaswamy*, Yale University*
Mark Leiserson, *University of Maryland*
Yue Li, *McGill University*
Jing Li, *Case Western Reserve University*
Kevin Liu, *Michigan State University*
Stefano Lonardi, *UC Riverside*

Ruibang Luo, *The University of Hong Kong*
Shaun Mahony, *The Pennsylvania State University*
Ion Mandoiu, *University of Connecticut*
Tijana Milenkovic, *University of Notre Dame*
Hosein Mohimani, *Carnegie Mellon University*
Ibrahim Numanagic, *Massachusetts Institute of Technology*
Robert Patro, *Stony Brook University*
Jian Peng, *University of Illinois at Urbana-Champaign*
Yann Ponty, *CNRS/LIX, Polytechnique*
Giuseppe Pozzi, *Politecnico di Milano*
Mattia Prosperi, *University of Florida*
Yanjun Qi, *University of Virginia*
Predrag Radivojac, *Northeastern University*
Sanguthevar Rajasekaran, *University of Connecticut*
Stephen Ramsey, *Oregon State University*
Sushmita Roy*, University of Wisconsin - Madison*
Leena Salmela*, University of Helsinki*
Mingfu Shao*, The Pennsylvania State University*
Amarda Shehu*, George Mason University*
Li Shen, *University of Pennsylvania*
Yang Shen, *Texas A&M University*
Xinghua Shi, *Temple University*
Ritambhara Singh, *University of Washington*
Paolo Soda, *Università Campus Bio-Medico di Roma*
Jens Stoye, *Bielefeld University*
Jimeng Sun, *Georgia Institute of Technology*
Wing-Kin Sung, *National University of Singapore*
Haixu Tang, *Indiana University Bloomington*
Tom Thorne, *University of Reading*
Giuseppe Tradigo, *University Magna Graecia of Catanzaro*
Ozlem Uzuner, *George Mason University*
Fabio Vandin, *University of Padova*
Jerome Waldispuhl, *McGill University*

May D. Wang, *Georgia Institute of Technology and Emory University*

Fei Wang, *Weill Cornell Medicine - Cornell University*

Haiying Wang, *University of Ulster*

Joshua Welch, *University of Michigan*

Yufeng Wu, *University of Connecticut*

Yonghui Wu, *University of Florida*

Jinbo Xu, *Toyota Technological Institute at Chicago*

Yuzhen Ye, *Indiana University Bloomington*

Kevin Yip, *Chinese University of Hong* Kong

Alex Zelikovsky, *Georgia State University*

Jianyang Zeng, *Tsinghua University*

Wei Zhang, *University of Central Florida*

Shihua Zhang, *Chinese Academy of Sciences*

Louxin Zhang, *National University of Singapore*

Degui Zhi, *University of Texas Health Science Center at Houston*

Shanfeng Zhu, *Fudan University*

Marinka Zitnik, *Stanford University*

Message from the General Chairs

It is our great pleasure to welcome you to the 10th ACM Conference on Bioinformatics, Computational Biology, and Health Informatics (ACM-BCB 2010). 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 42 peer-reviewed full research papers and 19 short research papers carefully selected from the 157 original submissions to the conference by diligent efforts of the program committee members. Additionally, we have 22 poster presentations and 5 highlight presentations on recently published, cutting-edge and high-impact research. The technical program is enriched by 7 pre-conference workshops and 7 tutorials. The program features diverse sessions including 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, Jian Ma and Pierangelo Veltri, as well as the army of program committee members who volunteered their time to rigorously review submissions. We thank Jianlin Cheng for selecting high-impact highlight presentations. Special thanks go to Wei Zhang and Xia Ning for putting together a strong and diverse workshop program, Shaun Mahony and Simina Boca for selecting interesting tutorials with broad appeal, and Dong Si and Daniel Mcskimming for bringing an outstanding set of timely research for display at the conference as poster presentations. We are thankful to Preetam Ghosh, yet again, handled registrations smoothly, and our proceedings chair Yang Shen, diligently and meticulously worked with Sheridan Communications and ACM headquarters to produce the conference proceedings in a timely manner for another year.

Special thanks go to Aidong Zhang, the Editor-in-Chief of IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB), and Dimitrios I. Fotiadis, the Editor-in-Chief of IEEE Journal of Biomedical and Health Informatics for approving ACM-BCB 2019 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 Robert (Bob) W. Cottingham for serving as the Demo and Exhibits Chair, and Pietro Hiram Guzzi, Paolo Soda, and Mattia Prosperi 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 Letu Qingge, Julien Kann, 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!

Xinghua (Mindy) Shi,

Temple University

Michael Buck,

University of Buffalo

Niagara Falls Conference & Event Center

First Street

Second Street

Rainbow Blvd.

Old Falls Street

Third Street

Loading Dock

Event Center
245' x 125' x 25'

Loading Dock

Kitchen

Green Room

Cascade Ballroom
130' x 80' x 16'

Grand Foyer
130' x 38' x 21'

Ticket Booth

Cataract
154' x 34'

Whitney
26' x 32'

Hennepin
26' x 25'

Adams
26' x 15'

Tubman
30' x 22'

Tesla
30' x 18'

Business Center

Olmsted
34' x 28'

DeVeaux
34' x 21'

Porter
34' x 31'

Red Jacket
28' x 40'

Governors Ampitheatre

Chairmans

Schoellkopf
34' x 26'

Taylor
24' x 16'

Hyde
4' x 15'

LaSalle
4' x 15'

Admin.

Castellani Museum

TGI Friday

N