

Preface.....v

PATTERN RECOGNITION IN BIOMEDICAL DATA: CHALLENGES IN PUTTING BIG DATA TO WORK

Session introduction.....1
Shefali Setia Verma, Anurag Verma, Dokyoon Kim, Christian Darabos

Learning Contextual Hierarchical Structure of Medical Concepts with Poincaré Embeddings to Clarify Phenotypes8
Brett K. Beaulieu-Jones, Isaac S. Kohane, Andrew L. Beam

The Effectiveness of Multitask Learning for Phenotyping with Electronic Health Records Data18
Daisy Yi Ding, Chloe Simpson, Stephen Pfohl, Dave C. Kale, Kenneth Jung, Nigam H. Shah

ODAL: A one-shot distributed algorithm to perform logistic regressions on electronic health records data from multiple clinical sites.....30
Rui Duan, Mary Regina Boland, Jason H. Moore, Yong Chen

PVC Detection Using a Convolutional Autoencoder and Random Forest Classifier42
Max Gordon, Cranos Williams

Removing Confounding Factors Associated Weights in Deep Neural Networks Improves the Prediction Accuracy for Healthcare Applications.....54
Haohan Wang, Zhenglin Wu, Eric P. Xing

DeepDom: Predicting protein domain boundary from sequence alone using stacked bidirectional LSTM.....66
Yuxu Jiang, Duolin Wang, Dong Xu

Res2s2aM: Deep residual network-based model for identifying functional noncoding SNPs in trait-associated regions.....76
Zheng Liu, Yao Yao, Qi Wei, Benjamin Weeder, Stephen A. Ramsey

DNA Steganalysis Using Deep Recurrent Neural Networks88
Ho Bae, Byunghan Lee, Sunyoung Kwon, Sungroh Yoon

Bi-directional Recurrent Neural Network Models for Geographic Location Extraction in Biomedical Literature100
Arjun Magge, Davy Weissenbacher, Abeed Sarker, Matthew Scotch, Graciela Gonzalez-Hernandez

Automatic Human-like Mining and Constructing Reliable Genetic Association Database with Deep Reinforcement Learning.....112
Haohan Wang, Xiang Liu, Yifeng Tao, Wenting Ye, Qiao Jin, William W. Cohen, Eric P. Xing

Estimating classification accuracy in positive-unlabeled learning: characterization and correction strategies124
Rashika Ramola, Shantanu Jain, Predrag Radivojac

<i>PLATYPUS: A Multiple-View Learning Predictive Framework for Cancer Drug Sensitivity Prediction</i>	136
Kiley Graim, Verena Friedl, Kathleen E. Houlahan, Joshua M. Stuart	
<i>Computational KIR copy number discovery reveals interaction between inhibitory receptor burden and survival</i>	148
Rachel M. Pyke, Raphael Genolet, Alexandre Harari, George Coukos, David Gfeller, Hannah Carter	
<i>Exploring microRNA Regulation of Cancer with Context-Aware Deep Cancer Classifier</i>	160
Blake Pyman, Alireza Sedghi, Shekoofeh Azizi, Kathrin Tyryshkin, Neil Renwick, Parvin Mousavi	
<i>Implementing and Evaluating A Gaussian Mixture Framework for Identifying Gene Function from TnSeq Data</i>	172
Kevin Li, Rachel Chen, William Lindsey, Aaron Best, Matthew DeJongh, Christopher Henry, Nathan Tintle	
<i>SNPs2ChIP: Latent Factors of ChIP-seq to infer functions of non-coding SNPs</i>	184
Shankara Anand, Laurynas Kalesinskas, Craig Smail, Yosuke Tanigawa	
<i>Extracting allelic read counts from 250,000 human sequencing runs in Sequence Read Archive</i>	196
Brian Tsui, Michelle Dow, Dylan Skola, Hannah Carter	
<i>Semantic workflows for benchmark challenges: Enhancing comparability, reusability and reproducibility</i>	208
Arunima Srivastava, Ravali Adusumilli, Hunter Boyce, Daniel Garijo, Varun Ratnakar, Rajiv Mayani, Thomas Yu, Raghu Machiraju, Yolanda Gil, Parag Mallick	
PRECISION MEDICINE: IMPROVING HEALTH THROUGH HIGH-RESOLUTION ANALYSIS OF PERSONAL DATA	
<i>Session introduction</i>	220
Steven E. Brenner, Martha Bulyk, Dana C. Crawford, Jill P. Mesirov, Alexander A. Morgan, Predrag Radivojac	
<i>CrowdVariant: a crowdsourcing approach to classify copy number variants</i>	224
Peyton Greenside, Justin Zook, Marc Salit, Madeleine Cule, Ryan Poplin, Mark DePristo	
<i>A repository of microbial marker genes related to human health and diseases for host phenotype prediction using microbiome data</i>	236
Wontack Han, Yuzhen Ye	
<i>AICM: A Genuine Framework for Correcting Inconsistency Between Large Pharmacogenomics Datasets</i>	248
Zhiyue Tom Hu, Yuting Ye, Patrick A. Newbury, Haiyan Huang, Bin Chen	

<i>Outgroup Machine Learning Approach Identifies Single Nucleotide Variants in Noncoding DNA Associated with Autism Spectrum Disorder</i>	260
Maya Varma, Kelley Marie Paskov, Jae-Yoon Jung, Brianna Sierra Chrisman, Nate Tyler Stockham, Peter Yigitcan Washington, Dennis Paul Wall	
<i>Detecting potential pleiotropy across cardiovascular and neurological diseases using univariate, bivariate, and multivariate methods on 43,870 individuals from the eMERGE network</i>	272
Xinyuan Zhang, Yogasudha Veturi, Shefali Verma, William Bone, Anurag Verma, Anastasia Lucas, Scott Hebring, Joshua C. Denny, Ian Stanaway, Gail P. Jarvik, David Crosslin, Eric B. Larson, Laura Rasmussen-Torvik, Sarah A. Pendergrass, Jordan W. Smoller, Hakon Hakonarson, Patrick Sleiman, Chunhua Weng, David Fasel, Wei-Qi Wei, Iftikhar Kullo, Daniel Schaid, Wendy K. Chung, Marylyn D. Ritchie	
<i>Integrating RNA expression and visual features for immune infiltrate prediction</i>	284
Derek Reiman, Lingdao Sha, Irvin Ho, Timothy Tan, Denise Lau, and Aly A. Khan	
<i>Influence of tissue context on gene prioritization for predicted transcriptome-wide association studies</i>	296
Binglan Li, Yogasudha Veturi, Yuki Bradford, Shefali S. Verma, Anurag Verma, Anastasia M. Lucas, David W. Haas, Marylyn D. Ritchie	
<i>Precision drug repurposing via convergent eQTL-based molecules and pathway targeting independent disease-associated polymorphisms</i>	308
Francesca Vitali, Joanne Berghout, Jungwei Fan, Jianrong Li, Qike Li, Haiquan Li, Yves A. Lussier	
<i>An Optimal Policy for Patient Laboratory Tests in Intensive Care Units</i>	320
Li-Fang Cheng, Niranjani Prasad, Barbara E Engelhardt	
SINGLE CELL ANALYSIS, WHAT IS IN THE FUTURE?	
<i>Session introduction</i>	332
Lana X. Garmire, Guo-Cheng Yuan, Rong Fan, Gene W. Yeo, John Quackenbush	
<i>LISA: Accurate reconstruction of cell trajectory and pseudo-time for massive single cell RNA-seq data</i>	338
Yang Chen, Yuping Zhang, Zhengqing Ouyang	
<i>Topological Methods for Visualization and Analysis of High Dimensional Single-Cell RNA Sequencing Data</i>	350
Tongxin Wang, Travis Johnson, Jie Zhang, Kun Huang	
<i>Parameter tuning is a key part of dimensionality reduction via deep variational autoencoders for single cell RNA transcriptomics</i>	362
Qiwen Hu, Casey S. Greene	
<i>Shallow Sparsely-Connected Autoencoders for Gene Set Projection</i>	374
Maxwell P. Gold, Alexander LeNail, and Ernest Fraenkel	

WHEN BIOLOGY GETS PERSONAL: HIDDEN CHALLENGES OF PRIVACY AND ETHICS IN BIOLOGICAL BIG DATA

<i>Session introduction</i>	386
Gamze Gürsoy, Arif Harmanci, Haixu Tang, Erman Ayday, Steven E. Brenner	
<i>Leveraging summary statistics to make inferences about complex phenotypes in large biobanks</i>	391
Angela Gasdaska, Derek Friend, Rachel Chen, Jason Westra, Matthew Zawistowski, William Lindsey, Nathan Tintle	
<i>Protecting Genomic Data Privacy with Probabilistic Modeling</i>	403
Sean Simmons, Bonnie Berger, Cenk Sahinalp	
<i>Evaluation of patient re-identification using laboratory test orders and mitigation via latent space variables</i>	415
Kipp W. Johnson, Jessica K. De Freitas, Benjamin S. Glicksberg, Jason R. Bobe, Joel T. Dudley	
<i>Implementing a universal informed consent process for the All of Us Research Program</i>	427
Megan Doerr, Shira Grayson, Sarah Moore, Christine Suver, John Wilbanks, Jennifer Wagner	

WORKSHOPS

<i>Merging heterogeneous clinical data to enable knowledge discovery</i>	439
Martin G. Seneviratne, Michael G. Kahn, Tina Hernandez-Boussard	
<i>Reading between the genes: interpreting non-coding DNA in high-throughput</i>	444
Joanne Berghout, Yves A. Lussier, Francesca Vitali, Martha L. Bulyk, Maricel G. Kann, Jason H. Moore	
<i>Text Mining and Machine Learning for Precision Medicine</i>	449
Graciela Gonzalez, Zhiyong Lu, Robert Leaman, Davy Weissenbacher, Mary Regina Boland, Yong Chen, Jingcheng Du, Juliane Fluck, Casey S. Greene, John Holmes, Aditya Kashyap, Rikke Linnemann Nielsen, Zhengqing Ouyang, Sebastian Schaaf, Jaclyn N. Taroni, Cui Tao, Yuping Zhang, Hongfang Liu,	
<i>Translational informatics of population Health: How large biomolecular and clinical datasets unite</i>	455
Yves A. Lussier, Atul Butte, Haiquan Li, Rong Chen, Jason H. Moore	