

Preface.....	vii
--------------	-----

## **APPLICATIONS OF GENETICS, GENOMICS AND BIOINFORMATICS IN DRUG DISCOVERY**

<i>Session introduction</i> .....	1
Richard Bourgon, Frederick E. Dewey, Zhengyan Kan, Shuyu D. Li	
<i>Characterization of drug-induced splicing complexity in prostate cancer cell line using long read technology</i> .....	8
Xintong Chen, Sander Houten, Kimaada Allette, Robert P. Sebra, Gustavo Stolovitzky, Bojan Losic	
<i>Prediction of protein-ligand interactions from paired protein sequence motifs and ligand substructures</i> .....	20
Peyton Greenside, Maureen Hillenmeyer, Anshul Kundaje	
<i>Cell-specific prediction and application of drug-induced gene expression profiles</i> .....	32
Rachel Hodos, Ping Zhang, Hao-Chih Lee, Qiaonan Duan, Zichen Wang, Neil R. Clark, Avi Ma'ayan, Fei Wang, Brian Kidd, Jianying Hu, David Sontag, Joel Dudley	
<i>Large-scale integration of heterogeneous pharmacogenomic data for identifying drug mechanism of action</i> .....	44
Yunan Luo, Sheng Wang, Jinfeng Xiao, Jian Peng	
<i>Chemical reaction vector embeddings: towards predicting drug metabolism in the human gut microbiome</i> .....	56
Emily K. Mallory, Ambika Acharya, Stefano E. Rensi, Peter J Turnbaugh, Roselie A. Bright, Russ B. Altman	
<i>Loss-of-function of neuroplasticity-related genes confers risk for human neurodevelopmental disorders</i> .....	68
Milo R. Smith, Benjamin S. Glicksberg, Li Li, Rong Chen, Hirofumi Morishita, Joel T. Dudley	
<i>Extracting a biologically relevant latent space from cancer transcriptomes with variational autoencoders</i> .....	80
Gregory P. Way, Casey S. Greene	
<i>Diffusion mapping of drug targets on disease signaling network elements reveals drug combination strategies</i> .....	92
Jielin Xu, Kelly Regan-Fendt, Siyuan Deng, William E. Carson III, Philip R.O. Payne, Fuhai Li	

## **CHALLENGES OF PATTERN RECOGNITION IN BIOMEDICAL DATA**

<i>Session introduction</i> .....	104
Shefali Setia Verma, Anurag Verma, Anna Okula Basile, Marta-Byrska Bishop, Christian Darabos	
<i>Large-scale analysis of disease pathways in the human interactome</i> .....	111
Monica Agrawal, Marinka Zitnik, Jure Leskovec	

<i>Mapping patient trajectories using longitudinal extraction and deep learning in the MIMIC-III Critical Care Database</i> .....	123
Brett K. Beaulieu-Jones, Patryk Orzechowski, Jason H. Moore	
<i>OWL-NETS: Transforming OWL representations for improved network inference</i> .....	134
Tiffany J. Callahan, William A. Baumgartner Jr., Michael Bada, Adrienne L. Stefanski, Ignacio Tripodi, Elizabeth K. White, Lawrence E. Hunter	
<i>Automated disease cohort selection using word embeddings from Electronic Health Records</i> .....	146
Benjamin S. Glicksberg, Riccardo Miotto, Kipp W. Johnson, Khader Shameer, Li Li, Rong Chen, Joel T. Dudley	
<i>Functional network community detection can disaggregate and filter multiple underlying pathways in enrichment analyses</i> .....	158
Lia X. Harrington, Gregory P. Way, Jennifer A. Doherty, Casey S. Greene	
<i>An ultra-fast and scalable quantification pipeline for transposable elements from next generation sequencing data</i> .....	169
Hyun-Hwan Jeong, Hari Krishna Yalamanchili, Caiwei Guo, Joshua M. Shulman, Zhandong Liu	
<i>Causal inference on electronic health records to assess blood pressure treatment targets: An application of the parametric g formula</i> .....	181
Kipp W. Johnson, Benjamin S. Glicksberg, Rachel Hodos, Khader Shameer, Joel T. Dudley	
<i>Data-driven advice for applying machine learning to bioinformatics problems</i> .....	193
Randal S. Olson, William La Cava, Zairah Mustahsan, Akshay Varik, Jason H. Moore	
<i>Improving the explainability of Random Forest classifier – user centered approach</i> .....	205
Dragutin Petkovic, Russ B. Altman, Mike Wong, Arthur Vigil	
<i>Tree-based methods for characterizing tumor density heterogeneity</i> .....	217
Katherine Shoemaker, Brian P. Hobbs, Karthik Bharath, Chaan S. Ng, Veerabhadran Baladandayuthapani	
<i>How powerful are summary-based methods for identifying expression-trait associations under different genetic architectures?</i> .....	229
Yogasudha Veturi, Marylyn D. Ritchie	

## **DEMOCRATIZING HEALTH DATA FOR TRANSLATIONAL RESEARCH**

<i>Session introduction</i> .....	241
Philip R.O Payne, Nigam H. Shah, Jessica D. Tenenbaum, Lara Mangravite	
<i>ClinGen Cancer Somatic Working Group – Standardizing and democratizing access to cancer molecular diagnostic data to drive translational research</i> .....	248
Subha Madhavan, Deborah Ritter, Christine Micheel, Shruti Rao, Angshumoy Roy, Dmitriy Sonkin, Matthew McCoy, Malachi Griffith, Obi L Griffith, Peter Mccarvey, Shashikant Kulkarni on Behalf of the ClinGen Somatic Working Group	

<i>A heuristic method for simulating open-data of arbitrary complexity that can be used to compare and evaluate machine learning methods.....</i>	260
Jason H. Moore, Maksim Shestov, Peter Schmitt, Randal S. Olson	
<i>Identifying natural health product and dietary supplement information within adverse event reporting systems .....</i>	269
Vivekanand Sharma, Indra Neil Sarkar	
<i>Best practices and lessons learned from reuse of 4 patient-derived metabolomics datasets in Alzheimer’s disease.....</i>	281
Jessica D. Tenenbaum, Colette Blach	
<i>Democratizing data science through data science training .....</i>	293
John Darrell Van Horn, Lily Fierro, Jeana Kamdar, Jonathan Gordon, Crystal Stewart, Avnish Bhattarai, Sumiko Abe, Xiaoxiao Lei, Caroline O’Driscoll, Aakanchha Sinha, Priyambada Jain, Gully Burns, Kristina Lerman, José Luis Ambite	

## **IMAGING GENOMICS**

<i>Session introduction.....</i>	305
Heng Huang, Li Shen, Paul M. Thompson, Kun Huang, Junzhou Huang, Lin Yang	
<i>Heritability estimates on resting state fMRI data using the ENIGMA analysis pipeline .....</i>	308
Bhim M. Adhikari, Neda Jahanshad, Dinesh Shukla, David C. Glahn, John Blangero, Richard C. Reynolds, Robert W. Cox, Els Fieremans, Jelle Veraart, Dmitry S. Novikov, Thomas E. Nichols, L. Elliot Hong, Paul M. Thompson, Peter Kochunov	
<i>Discriminative bag-of-cells for imaging-genomics .....</i>	320
Benjamin Chidester, Minh N. Do, Jian Ma	
<i>MRI to MGMT: Predicting methylation status in glioblastoma patients using convolutional recurrent neural networks.....</i>	332
Lichy Han, Maulik R. Kamdar	
<i>Deep integrative analysis for survival prediction.....</i>	344
Chenglong Huang, Albert Zhang, Guanghua Xiao	
<i>Genotype-Phenotype association study via new multi-task learning model.....</i>	354
Zhouyuan Huo, Dinggang Shen, Heng Huang	
<i>Codon bias among synonymous rare variants is associated with Alzheimer’s disease imaging biomarker.....</i>	366
Jason E. Miller, Manu K. Shivakumar, Shannon L. Risacher, Andrew J. Saykin, Seunggeun Lee, Kwangsik Nho, Dokyoon Kim, for the Alzheimer’s Disease Neuroimaging Initiative (ADNI)	
<i>Building trans-omics evidence: using imaging and ‘omics’ to characterize cancer profiles.....</i>	378
Arunima Srivastava, Chaitanya Kulkarni, Parag Mallick, Kun Huang, Raghu Machiraju	

## **PRECISION MEDICINE: FROM DIPTYPES TO DISPARITIES TOWARDS**

## IMPROVED HEALTH AND THERAPIES

<i>Session Introduction</i> .....	390
Dana C. Crawford, Alexander A. Morgan, Joshua C. Denny, Bruce J. Aronow, Steven E. Brenner	
<i>Single subject transcriptome analysis to identify functionally signed gene set or pathway activity</i> .....	401
Joanne Berghout, Qike Li, Nima Pouladi, Jianrong Li, Yves A. Lussier	
<i>Using simulation and optimization approach to improve outcome through warfarin precision treatment</i> .....	413
Chih-Lin Chi, Lu He, Kourosh Ravvaz, John Weissert, Peter J. Tonellato	
<i>Local ancestry transitions modify snp-trait associations</i> .....	425
Alexandra E. Fish, Dana C. Crawford, John A. Capra, William S. Bush	
<i>Coalitional game theory as a promising approach to identify candidate autism genes</i> .....	437
Anika Gupta, Min Woo Sun, Kelley M. Paskov, Nate T. Stockham, Jae-Yoon Jung, Dennis P. Wall	
<i>Evaluation of PrediXcan for prioritizing GWAS associations and predicting gene expression</i> .....	449
Binglan Li, Shefali S. Verma, Yogasudha C. Veturi, Anurag Verma, Yuki Bradford, David W. Haas, Marylyn D. Ritchie	
<i>Considerations for automated machine learning in clinical metabolic profiling: Altered homocysteine plasma concentration associated with metformin exposure</i> .....	461
Alena Orlenko, Jason H. Moore, Patryk Orzechowski, Randal S. Olson, Junmei Cairns, Pedro J. Caraballo, Richard M. Weinshilboum, Liewei Wang, Matthew K. Breitenstein	
<i>Addressing vital sign alarm fatigue using personalized alarm thresholds</i> .....	473
Sarah Poole, Nigam Shah	
<i>Emergence of pathway-level composite biomarkers from converging gene set signals of heterogeneous transcriptomic responses</i> .....	485
Samir Rachid Zaim, Qike Li, A. Grant Schissler, Yves A. Lussier	
<i>Analyzing metabolomics data for association with genotypes using two-component Gaussian mixture distributions</i> .....	497
Jason Westra, Nicholas Hartman, Bethany Lake, Gregory Shearer, Nathan Tintle	

## READING BETWEEN THE GENES: COMPUTATIONAL MODELS TO DISCOVER FUNCTION FROM NONCODING DNA

<i>Session Introduction</i> .....	508
Yves A. Lussier†, Joanne Berghout, Francesca Vitali, Kenneth S. Ramos, Maricel Kann, Jason H. Moore	
<i>Pan-cancer analysis of expressed somatic nucleotide variants in long intergenic non-coding RNA</i> .....	513
Travers Ching, Lana X. Garmire	

<i>Convergent downstream candidate mechanisms of independent intergenic polymorphisms between co-classified diseases implicate epistasis among noncoding elements.....</i>	525
Jiali Han, Jianrong Li, Ikbel Achour, Lorenzo Pesce, Ian Foster, Haiquan Li, Yves A. Lussier	
<i>Network analysis of pseudogene-gene relationships: from pseudogene evolution to their functional potentials.....</i>	537
Travis S. Johnson, Sihong Li, Jonathan R. Kho, Kun Huang, Yan Zhang	
<i>Leveraging putative enhancer-promoter interactions to investigate two-way epistasis in Type 2 Diabetes GWAS.....</i>	549
Elisabetta Manduchi, Alessandra Chesi, Molly A. Hall, Struan F. A. Grant, Jason H. Moore	

## **ADVANCES IN TEXT MINING AND VISUALIZATION FOR PRECISION MEDICINE**

<i>Session Introduction.....</i>	560
Graciela Gonzalez-Hernandez, Abeed Sarker, Karen O'Connor, Casey Greene, Hongfang Liu	
<i>Improving precision in concept normalization.....</i>	567
Mayla Boguslav, K. Bretonnel Cohen, William A. Baumgartner Jr., Lawrence E. Hunter	
<i>VisAGE: Integrating external knowledge into electronic medical record visualization.....</i>	579
Edward W. Huang, Sheng Wang, ChengXiang Zhai	
<i>GeneDive: A gene interaction search and visualization tool to facilitate precision medicine.....</i>	591
Paul Previde, Brook Thomas, Mike Wong, Emily K. Mallory, Dragutin Petkovic, Russ B. Altman, Anagha Kulkarni	
<i>Annotating gene sets by mining large literature collections with protein networks.....</i>	603
Sheng Wang, Jianzhu Ma, Michael Ku Yu, Fan Zheng, Edward W. Huang, Jiawei Han, Jian Peng, Trey Ideker	

## **WORKSHOPS**

<i>The diversity and disparity in biomedical informatics (DDBI) workshop.....</i>	615
William Southerland, S. Joshua Swamidass, Philip R. O. Payne, Laura Wiley, ClarLynda Williams-DeVane	
<i>Integrating community-level data resources for precision medicine research.....</i>	619
William S. Bush, Dana C. Crawford, Farren Briggs, Darcy Freedman	
<i>Machine learning and deep analytics for biocomputing: Call for better explainability.....</i>	624
Dragutin Petkovic, Lester Kobzik, Christopher Re	
<i>Methods for examining data quality in healthcare integrated data repositories.....</i>	629
Vojtech Huser, Michael G. Kahn, Jeffrey S. Brown, Ramkiran Gouripeddi	

## ERRATUM

<i>ERRATUM: Identifying mutation specific cancer pathways using a structurally resolved protein interaction network</i> .....	635
H. Billur Engin, Matan Hofree, Hannah Carter	