

Pacific Symposium on Biocomputing 2018

Fairmont Orchid, Hawaii, USA
3-7 January 2018

ISBN: 978-1-7138-2794-8

Printed from e-media with permission by:

Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571



Some format issues inherent in the e-media version may also appear in this print version.

The full collection of papers in these proceedings is licensed under a Creative Commons Attribution 4.0 International License.

License details: <http://creativecommons.org/licenses/by/4.0/>.

The individual papers are © 2017 The Authors. Distributed under the terms of the Creative Commons Attribution Non-Commercial (CC BY-NC) 4.0 License

License details: <https://creativecommons.org/licenses/by-nc/4.0/>.

Printed with permission by Curran Associates, Inc. (2021)

World Scientific Publishing Co., Inc.
27 Warren Street, Suite 401-402
Hackensack, NJ 07601

Phone: 1-201-487-9655

Fax: 1-201-487-9656

www.worldscientific.com

Additional copies of this publication are available from:

Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571 USA
Phone: 845-758-0400
Fax: 845-758-2634
Email: curran@proceedings.com
Web: www.proceedings.com

TABLE OF CONTENTS

APPLICATIONS OF GENETICS, GENOMICS AND BIOINFORMATICS IN DRUG DISCOVERY

SESSION INTRODUCTION - APPLICATIONS OF GENETICS, GENOMICS AND BIOINFORMATICS IN DRUG DISCOVERY 1
Richard Bourgon, Frederick E. Dewey, Zhengyan Kan, Shuyu D. Li

CHARACTERIZATION OF DRUG-INDUCED SPLICING COMPLEXITY IN PROSTATE CANCER CELL LINE USING LONG READ TECHNOLOGY 8
Xintong Chen, Sander Houten, Kimaada Allette, Robert P. Sebra, Gustavo Stolovitzky, Bojan Losic

PREDICTION OF PROTEIN-LIGAND INTERACTIONS FROM PAIRED PROTEIN SEQUENCE MOTIFS AND LIGAND SUBSTRUCTURES 20
Peyton Greenside, Maureen Hillenmeyer, Anshul Kundaje

CELL-SPECIFIC PREDICTION AND APPLICATION OF DRUG-INDUCED GENE EXPRESSION 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

LARGE-SCALE INTEGRATION OF HETEROGENEOUS PHARMACOGENOMIC DATA FOR IDENTIFYING DRUG MECHANISM OF ACTION 44
Yunan Luo, Sheng Wang, Jinfeng Xiao, Jian Peng

CHEMICAL REACTION VECTOR EMBEDDINGS: TOWARDS PREDICTING DRUG METABOLISM IN THE HUMAN GUT MICROBIOME 56
Emily K. Mallory, Ambika Acharya, Stefano E. Rensi, Peter J Turnbaugh, Roselie A. Bright, Russ B. Altman

LOSS-OF-FUNCTION OF NEUROPLASTICITY-RELATED GENES CONFERS RISK FOR HUMAN NEURODEVELOPMENTAL DISORDERS 68
Milo R. Smith, Benjamin S. Glicksberg, Li Li, Rong Chen, Hirofumi Morishita, Joel T. Dudley

EXTRACTING A BIOLOGICALLY RELEVANT LATENT SPACE FROM CANCER TRANSCRIPTOMES WITH VARIATIONAL AUTOENCODERS 80
Gregory P. Way, Casey S. Greene

DIFFUSION MAPPING OF DRUG TARGETS ON DISEASE SIGNALING NETWORK ELEMENTS REVEALS DRUG COMBINATION STRATEGIES 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

SESSION INTRODUCTION - CHALLENGES OF PATTERN RECOGNITION IN BIOMEDICAL DATA 104
Shefali Setia Verma, Anurag Verma, Anna Okula Basile, Marta-Byrska Bishop, Christian Darabos

LARGE-SCALE ANALYSIS OF DISEASE PATHWAYS IN THE HUMAN INTERACTOME	111
<i>Monica Agrawal, Marinka Zitnik, Jure Leskovec</i>	
MAPPING PATIENT TRAJECTORIES USING LONGITUDINAL EXTRACTION AND DEEP LEARNING IN THE MIMIC-III CRITICAL CARE DATABASE	123
<i>Brett K. Beaulieu-Jones, Patryk Orzechowski, Jason H. Moore</i>	
OWL-NETS: TRANSFORMING OWL REPRESENTATIONS FOR IMPROVED NETWORK INFERENCE	133
<i>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	145
<i>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	157
<i>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	168
<i>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	180
<i>Kipp W. Johnson, Benjamin S. Glicksberg, Rachel Hodos, Khader Shameer, Joel T. Dudley</i>	
DATA-DRIVEN ADVICE FOR APPLYING MACHINE LEARNING TO BIOINFORMATICS PROBLEMS	192
<i>Randal S. Olson, William La Cava, Zairah Mustahsan, Akshay Varik, Jason H. Moore</i>	
IMPROVING THE EXPLAINABILITY OF RANDOM FOREST CLASSIFIER - USER CENTERED APPROACH	204
<i>Dragutin Petkovic, Russ B. Altman, Mike Wong, Arthur Vigil</i>	
TREE-BASED METHODS FOR CHARACTERIZING TUMOR DENSITY HETEROGENEITY	216
<i>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?	228
<i>Yogasudha Veturi, Marylyn D. Ritchie</i>	
<u>DEMOCRATIZING HEALTH DATA FOR TRANSLATIONAL RESEARCH</u>	
SESSION INTRODUCTION - DEMOCRATIZING HEALTH DATA FOR TRANSLATIONAL RESEARCH	240
<i>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.....	247
<i>Subha Madhavan, Deborah Ritter, Christine Micheel, Shruti Rao, Angshumoy Roy, Dmitriy Sonkin, Matthew McCoy, Malachi Griffith, Obi L Griffith, Peter McGarvey, Shashikant Kulkarni</i>	
A HEURISTIC METHOD FOR SIMULATING OPEN-DATA OF ARBITRARY COMPLEXITY THAT CAN BE USED TO COMPARE AND EVALUATE MACHINE LEARNING METHODS	259
<i>Jason H. Moore, Maksim Shestov, Peter Schmitt, Randal S. Olson</i>	
IDENTIFYING NATURAL HEALTH PRODUCT AND DIETARY SUPPLEMENT INFORMATION WITHIN ADVERSE EVENT REPORTING SYSTEMS	268
<i>Vivekanand Sharma, Indra Neil Sarkar</i>	
BEST PRACTICES AND LESSONS LEARNED FROM REUSE OF 4 PATIENT-DERIVED METABOLOMICS DATASETS IN ALZHEIMER'S DISEASE	280
<i>Jessica D. Tenenbaum, Colette Blach</i>	
DEMOCRATIZING DATA SCIENCE THROUGH DATA SCIENCE TRAINING	292
<i>John Darrell Van Horn, Lily Fierro, Jeana Kamdar, Jonathan Gordon, Crystal Stewart, Avnish Bhattra, Sumiko Abe, Xiaoxiao Lei, Caroline O'Driscoll, Aakanchha Sinha, Priyambada Jain, Gully Burns, Kristina Lerman, Jose Luis Ambite</i>	
SESSION INTRODUCTION - IMAGING GENOMICS	304
<i>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	307
<i>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.....	319
<i>Benjamin Chidester, Minh N. Do, Jian Ma</i>	
MRI TO MGMT: PREDICTING METHYLATION STATUS IN GLIOBLASTOMA PATIENTS USING CONVOLUTIONAL RECURRENT NEURAL NETWORKS	331
<i>Lichy Han, Maulik R. Kamdar</i>	
DEEP INTEGRATIVE ANALYSIS FOR SURVIVAL PREDICTION	343
<i>Chenglong Huang, Albert Zhang, Guanghua Xiao</i>	
GENOTYPE-PHENOTYPE ASSOCIATION STUDY VIA NEW MULTI-TASK LEARNING MODEL.....	353
<i>Zhouyuan Huo, Dinggang Shen, Heng Huang</i>	
CODON BIAS AMONG SYNONYMOUS RARE VARIANTS IS ASSOCIATED WITH ALZHEIMER'S DISEASE IMAGING BIOMARKER	365
<i>Jason E. Miller, Manu K. Shivakumar, Shannon L. Risacher, Andrew J. Saykin, Seunggeun Lee, Kwangsik Nho, Dokyoon Kim</i>	
BUILDING TRANS-OMICS EVIDENCE: USING IMAGING AND 'OMICS' TO CHARACTERIZE CANCER PROFILES	377
<i>Arunima Srivastava, Chaitanya Kulkarni, Parag Mallick, Kun Huang, Raghu Machiraju</i>	

PRECISION MEDICINE: FROM DILOTYPES TO DISPARITIES TOWARDS IMPROVED HEALTH AND THERAPIES

SESSION INTRODUCTION - PRECISION MEDICINE: FROM DILOTYPES TO DISPARITIES TOWARDS IMPROVED HEALTH AND THERAPIES 389
Dana C. Crawford, Alexander A. Morgan, Joshua C. Denny, Bruce J. Aronow, Steven E. Brenner

SINGLE SUBJECT TRANSCRIPTOME ANALYSIS TO IDENTIFY FUNCTIONALLY SIGNED GENE SET OR PATHWAY ACTIVITY..... 400
Joanne Berghout, Qike Li, Nima Pouladi, Jianrong Li, Yves A. Lussier

USING SIMULATION AND OPTIMIZATION APPROACH TO IMPROVE OUTCOME THROUGH WARFARIN PRECISION TREATMENT..... 412
Chih-Lin Chi, Lu He, Kouros Ravvaz, John Weissert, Peter J. Tonellato

LOCAL ANCESTRY TRANSITIONS MODIFY SNP-TRAIT ASSOCIATIONS 424
Alexandra E. Fish, Dana C. Crawford, John A. Capra, William S. Bush

COALITIONAL GAME THEORY AS A PROMISING APPROACH TO IDENTIFY CANDIDATE AUTISM GENES..... 436
Anika Gupta, Min Woo Sun, Kelley M. Paskov, Nate T. Stockham, Jae-Yoon Jung, Dennis P. Wall

EVALUATION OF PREDIXCAN FOR PRIORITIZING GWAS ASSOCIATIONS AND PREDICTING GENE EXPRESSION..... 448
Binglan Li, Shefali S. Verma, Yogasudha C. Veturi, Anurag Verma, Yuki Bradford, David W. Haas, Marylyn D. Ritchie

CONSIDERATIONS FOR AUTOMATED MACHINE LEARNING IN CLINICAL METABOLIC PROFILING: ALTERED HOMOCYSTEINE PLASMA CONCENTRATION ASSOCIATED WITH METFORMIN EXPOSURE 460
Alena Orlenko, Jason H. Moore, Patryk Orzechowski, Randal S. Olson, Junmei Cairns, Pedro J. Caraballo, Richard M. Weinshilboun, Liewei Wang, Matthew K. Breitenstein

ADDRESSING VITAL SIGN ALARM FATIGUE USING PERSONALIZED ALARM THRESHOLDS 472
Sarah Poole, Nigam Shah

EMERGENCE OF PATHWAY-LEVEL COMPOSITE BIOMARKERS FROM CONVERGING GENE SET SIGNALS OF HETEROGENEOUS TRANSCRIPTOMIC RESPONSES..... 484
Samir Rachid Zaim, Qike Li, A. Grant Schissler, Yves A. Lussier

ANALYZING METABOLOMICS DATA FOR ASSOCIATION WITH GENOTYPES USING TWO-COMPONENT GAUSSIAN MIXTURE DISTRIBUTIONS..... 496
Jason Westra, Nicholas Hartman, Bethany Lake, Gregory Shearer, Nathan Tintle

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

SESSION INTRODUCTION - READING BETWEEN THE GENES: COMPUTATIONAL MODELS TO DISCOVER FUNCTION FROM NONCODING DNA 507
Yves A. Lussier, Joanne Berghout, Francesca Vitali, Kenneth S. Ramos, Maricel Kann, Jason H. Moore

PAN-CANCER ANALYSIS OF EXPRESSED SOMATIC NUCLEOTIDE VARIANTS IN LONG INTERGENIC NON-CODING RNA	512
<i>Travers Ching, Lana X. Garmire</i>	
CONVERGENT DOWNSTREAM CANDIDATE MECHANISMS OF INDEPENDENT INTERGENIC POLYMORPHISMS BETWEEN CO-CLASSIFIED DISEASES IMPLICATE EPISTASIS AMONG NONCODING ELEMENTS	524
<i>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.....	536
<i>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.....	548
<i>Elisabetta Manduchi, Alessandra Chesi, Molly A. Hall, Struan F. A. Grant, Jason H. Moore</i>	

ADVANCES IN TEXT MINING AND VISUALIZATION FOR PRECISION MEDICINE

SESSION INTRODUCTION - ADVANCES IN TEXT MINING AND VISUALIZATION FOR PRECISION MEDICINE.....	559
<i>Graciela Gonzalez-Hernandez, Abeed Sarker, Karen O'Connor, Casey Greene, Hongfang Liu</i>	
IMPROVING PRECISION IN CONCEPT NORMALIZATION	566
<i>Mayla Boguslav, K. Bretonnel Cohen, William A. Baumgartner Jr., Lawrence E. Hunter</i>	
VISAGE: INTEGRATING EXTERNAL KNOWLEDGE INTO ELECTRONIC MEDICAL RECORD VISUALIZATION	578
<i>Edward W. Huang, Sheng Wang, Chengxiang Zhai</i>	
GENEDIVE: A GENE INTERACTION SEARCH AND VISUALIZATION TOOL TO FACILITATE PRECISION MEDICINE	590
<i>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	602
<i>Sheng Wang, Jianzhu Ma, Michael Ku Yu, Fan Zheng, Edward W. Huang, Jiawei Han, Jian Peng, Trey Ideker</i>	

WORKSHOPS

THE DIVERSITY AND DISPARITY IN BIOMEDICAL INFORMATICS (DBBI) WORKSHOP	614
<i>William Southerland, S. Joshua Swamidass, Philip R. O. Payne, Laura Wiley, Clarlynda Williams-Devane</i>	
INTEGRATING COMMUNITY-LEVEL DATA RESOURCES FOR PRECISION MEDICINE RESEARCH	618
<i>William S. Bush, Dana C. Crawford, Farren Briggs, Darcy Freedman, Chantel Sloan</i>	
MACHINE LEARNING AND DEEP ANALYTICS FOR BIOCOMPUTING: CALL FOR BETTER EXPLAINABILITY	623
<i>Dragutin Petkovic, Lester Kobzik, Christopher Re</i>	

METHODS FOR EXAMINING DATA QUALITY IN HEALTHCARE INTEGRATED DATA REPOSITORIES	628
<i>Vojtech Huser, Michael G. Kahn, Jeffrey S. Brown, Ramkiran Gouripeddi</i>	

ERRATUM

ERRATUM: IDENTIFYING MUTATION SPECIFIC CANCER PATHWAYS USING A STRUCTURALLY RESOLVED PROTEIN INTERACTION NETWORK.....	634
<i>H. Billur Engin, Matan Hofree, Hannah Carter</i>	

Author Index