

Pacific Symposium on Biocomputing 2019

Fairmont Orchid, Hawaii, USA
3 - 7 January 2019

ISBN: 978-1-7138-2795-5

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 © 2020 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

PATTERN RECOGNITION IN BIOMEDICAL DATA: CHALLENGES IN PUTTING BIG DATA TO WORK	1
<i>S.S. Verma, A. Verma, D. Kim, C. Darabos</i>	
LEARNING CONTEXTUAL HIERARCHICAL STRUCTURE OF MEDICAL CONCEPTS WITH POINCAIRE EMBEDDINGS TO CLARIFY PHENOTYPES.....	8
<i>B.K. Beaulieu-Jones, I.S. Kohane, A.L. Beam</i>	
THE EFFECTIVENESS OF MULTITASK LEARNING FOR PHENOTYPING WITH ELECTRONIC HEALTH RECORDS DATA	18
<i>D.Y. Ding, C. Simpson, S. Pfohl, D.C. Kale, K. Jung, N.H. Shah</i>	
ODAL: A ONE-SHOT DISTRIBUTED ALGORITHM TO PERFORM LOGISTIC REGRESSIONS ON ELECTRONIC HEALTH RECORDS DATA FROM MULTIPLE CLINICAL SITES.....	30
<i>R. Duan, M.R. Boland, J.H. Moore, Y. Chen</i>	
PVC DETECTION USING A CONVOLUTIONAL AUTOENCODER AND RANDOM FOREST CLASSIFIER.....	42
<i>M. Gordon, C. Williams</i>	
REMOVING CONFOUNDING FACTORS ASSOCIATED WEIGHTS IN DEEP NEURAL NETWORKS IMPROVES THE PREDICTION ACCURACY FOR HEALTHCARE APPLICATIONS.....	54
<i>H. Wang, Z. Wu, E.P. Xing</i>	
DEEPPOM: PREDICTING PROTEIN DOMAIN BOUNDARY FROM SEQUENCE ALONE USING STACKED BIDIRECTIONAL LSTM	66
<i>Y. Jiang, D. Wang, D. Xu</i>	
RES2S2AM: DEEP RESIDUAL NETWORK-BASED MODEL FOR IDENTIFYING FUNCTIONAL NONCODING SNPS IN TRAIT-ASSOCIATED REGIONS.....	76
<i>Z. Liu, Y. Yao, Q. Wei, B. Weeder, S.A. Ramsey</i>	
DNA STEGANALYSIS USING DEEP RECURRENT NEURAL NETWORKS.....	88
<i>H. Bae, B. Lee, S. Kwon, S. Yoon</i>	
BI-DIRECTIONAL RECURRENT NEURAL NETWORK MODELS FOR GEOGRAPHIC LOCATION EXTRACTION IN BIOMEDICAL LITERATURE.....	100
<i>A. Magge, D. Weissenbacher, A. Sarker, M. Scotch, G. Gonzalez-Hernandez</i>	
AUTOMATIC HUMAN-LIKE MINING AND CONSTRUCTING RELIABLE GENETIC ASSOCIATION DATABASE WITH DEEP REINFORCEMENT LEARNING	112
<i>H. Wang, X. Liu, Y. Tao, W. Ye, Q. Jin, W.W. Cohen, E.P. Xing</i>	
ESTIMATING CLASSIFICATION ACCURACY IN POSITIVE-UNLABELED LEARNING: CHARACTERIZATION AND CORRECTION STRATEGIES	124
<i>R. Ramola, S. Jain, P. Radivojac</i>	
PLATYPUS: A MULTIPLE VIEW LEARNING PREDICTIVE FRAMEWORK FOR CANCER DRUG SENSITIVITY PREDICTION.....	136
<i>K. Graim, V. Friedl, K.E. Houlahan, J.M. Stuart</i>	

COMPUTATIONAL KIR COPY NUMBER DISCOVERY REVEALS INTERACTION BETWEEN INHIBITORY RECEPTOR BURDEN AND SURVIVAL	148
<i>R.M. Pyke, R. Genolet, A. Harari, G. Coukos, D. Gfeller, H. Carter</i>	
EXPLORING MICRORNA REGULATION OF CANCER WITH CONTEXT-AWARE DEEP CANCER CLASSIER.....	160
<i>B. Pyman, A. Sedghi, S. Aziz, K. Tyryshkin, N. Renwick, P. Mousavi</i>	
IMPLEMENTING AND EVALUATING A GAUSSIAN MIXTURE FRAMEWORK FOR IDENTIFYING GENE FUNCTION FROM TNSEQ DATA.....	172
<i>K. Li, R. Chen, W. Lindsey, A. Best, M. DeJongh, C. Henry, N. Tintle</i>	
SNPS2CHIP: LATENT FACTORS OF CHIP-SEQ TO INFER FUNCTIONS OF NON-CODING SNPS	184
<i>S. Anand, L. Kalesinskas, C. Smail, Y. Tanigawa</i>	
EXTRACTING ALLELIC READ COUNTS FROM 250,000 HUMAN SEQUENCING RUNS IN SEQUENCE READ ARCHIVE.....	196
<i>B. Tsui, M. Dow, D. Skola, H. Carter</i>	
SEMANTIC WORKFLOWS FOR BENCHMARK CHALLENGES: ENHANCING COMPARABILITY, REUSABILITY AND REPRODUCIBILITY	208
<i>A. Srivastava, R. Adusumilli, H. Boyce, D. Garijo, V. Ratnakar, R. Mayani, T. Yu, R. Machiraju, Y. Gil, P. Mallick</i>	
PRECISION MEDICINE: IMPROVING HEALTH THROUGH HIGH-RESOLUTION ANALYSIS OF PERSONAL DATA.....	220
<i>S.E. Brenner, M. Bulyk, D.C. Crawford, J.P. Mesirov, A.A. Morgan, P. Radivojac</i>	
CROWDVARIANT: A CROWDSOURCING APPROACH TO CLASSIFY COPY NUMBER VARIANTS.....	224
<i>P. Greenside, J. Zook, M. Salit, M. Cule, R. Poplin, M. DePristo</i>	
A REPOSITORY OF MICROBIAL MARKER GENES RELATED TO HUMAN HEALTH AND DISEASES FOR HOST PHENOTYPE PREDICTION USING MICROBIOME DATA.....	236
<i>W. Han, Y. Ye</i>	
AICM: A GENUINE FRAMEWORK FOR CORRECTING INCONSISTENCY BETWEEN LARGE PHARMACOGENOMICS DATASETS	248
<i>Z.T. Hu, Y. Ye, P.A. Newbury, H. Huang, B. Chen</i>	
OUTGROUP MACHINE LEARNING APPROACH IDENTIFIES SINGLE NUCLEOTIDE VARIANTS IN NONCODING DNA ASSOCIATED WITH AUTISM SPECTRUM DISORDER	260
<i>M. Varma, K.M. Paskov, J.Y. Jung, B.S. Chrisman, N.T. Stockham, P.Y. Washington, D.P. Wall</i>	
DETECTING POTENTIAL PLEIOTROPY ACROSS CARDIOVASCULAR AND NEUROLOGICAL DISEASES USING UNIVARIATE, BIVARIATE, AND MULTIVARIATE METHODS ON 43,870 INDIVIDUALS FROM THE EMERGE NETWORK	272
<i>X. Zhang, Y. Veturi, S. Verma, W. Bone, A. Verma, A. Lucas, S. Hebbbring, J.C. Denny, I.B. Stanaway, G.P. Jarvik, D. Crosslin, E.B. Larson, L. Rasmussen-Torvik, S.A. Pendergrass, J.W. Smoller, H. Hakonarson, P. Sleiman, C. Weng, D. Fasel, W.Q. Wei, I. Kullo, D. Schaid, W.K. Chung, M.D. Ritchie</i>	
INTEGRATING RNA EXPRESSION AND VISUAL FEATURES FOR IMMUNE INFILTRATE PREDICTION	284
<i>D. Reiman, L. Sha, I. Ho, T. Tan, D. Lau, A.A. Khan</i>	

INFLUENCE OF TISSUE CONTEXT ON GENE PRIORITIZATION FOR PREDICTED TRANSCRIPTOME-WIDE ASSOCIATION STUDIES	296
<i>B. Li, Y. Veturi, Y. Bradford, S.S. Verma, A. Verma, A.M. Lucas, D.W. Haas, M.D. Ritchie</i>	
PRECISION DRUG REPURPOSING VIA CONVERGENT EQTL-BASED MOLECULES AND PATHWAY TARGETING INDEPENDENT DISEASE-ASSOCIATED POLYMORPHISMS.....	308
<i>F. Vitali, J. Berghout, J. Fan, J. Li, Q. Li, H. Li, Y.A. Lussier</i>	
AN OPTIMAL POLICY FOR PATIENT LABORATORY TESTS IN INTENSIVE CARE UNITS	320
<i>L.F. Cheng, N. Prasad, B.E. Engelhardt</i>	
SINGLE CELL ANALYSIS, WHAT IS IN THE FUTURE?.....	332
<i>L.X. Garmire, G.C. Yuan, R. Fan, G.W. Yeo, J. Quackenbush</i>	
LISA: ACCURATE RECONSTRUCTION OF CELL TRAJECTORY AND PSEUDO-TIME FOR MASSIVE SINGLE CELL RNA-SEQ DATA	338
<i>Y. Chen, Y. Zhang, Z. Ouyang</i>	
TOPOLOGICAL METHODS FOR VISUALIZATION AND ANALYSIS OF HIGH DIMENSIONAL SINGLE-CELL RNA SEQUENCING DATA	350
<i>T. Wang, T. Johnson, J. Zhang, K. Huang</i>	
PARAMETER TUNING IS A KEY PART OF DIMENSIONALITY REDUCTION VIA DEEP VARIATIONAL AUTOENCODERS FOR SINGLE CELL RNA TRANSCRIPTOMICS.....	362
<i>Q. Hu, C.S. Greene</i>	
SHALLOW SPARSELY-CONNECTED AUTOENCODERS FOR GENE SET PROJECTION	374
<i>M.P. Gold, A. LeNail, E. Fraenkel</i>	
WHEN BIOLOGY GETS PERSONAL: HIDDEN CHALLENGES OF PRIVACY AND ETHICS IN BIOLOGICAL BIG DATA.....	386
<i>G. Gursoy, A. Harmanaci, H. Tang, E. Ayday, S.E. Brenner</i>	
LEVERAGING SUMMARY STATISTICS TO MAKE INFERENCES ABOUT COMPLEX PHENOTYPES IN LARGE BIOBANKS.....	391
<i>A. Gasdaska, D. Friend, R. Chen, J. Westra, M. Zawistowski, W. Lindsey, N. Tintle</i>	
PROTECTING GENOMIC DATA PRIVACY WITH PROBABILISTIC MODELING	403
<i>S. Simmons, B. Berger, C. Sahinalp</i>	
EVALUATION OF PATIENT RE-IDENTIFICATION USING LABORATORY TEST ORDERS AND MITIGATION VIA LATENT SPACE VARIABLES	415
<i>K.W. Johnson, J.K. De Freitas, B.S. Glicksberg, J.R. Bobe, J.T. Dudley</i>	
IMPLEMENTING A UNIVERSAL INFORMED CONSENT PROCESS FOR THE ALL OF US RESEARCH PROGRAM	427
<i>M. Doerr, S. Grayson, S. Moore, C. Suver, J. Wilbanks, J. Wagner</i>	
MERGING HETEROGENEOUS CLINICAL DATA TO ENABLE KNOWLEDGE DISCOVERY	439
<i>M.G. Seneviratne, M.G. Kahn, T. Hernandez-Boussard</i>	
WORKSHOP DURING THE PACIFIC SYMPOSIUM OF BIOCOMPUTING, JAN 3-7, 2019: READING BETWEEN THE GENES: INTERPRETING NON-CODING DNA IN HIGH- THROUGHPUT	444
<i>J. Berghout, Y.A. Lussier, F. Vitali, M.L. Bulyk, M.G. Kann, J.H. Moore</i>	

PSB 2019 WORKSHOP ON TEXT MINING AND VISUALIZATION FOR PRECISION
MEDICINE 449
*G. Gonzalez-Hernandez, Z. Lu, R. Leaman, D. Weissenbacher, M.R. Boland, Y. Chen, J. Du,
J. Fluck, C.S. Greene, J. Holmes, A. Kashyap, R.L. Nielsen, Z. Ouyang, S. Schaaf, J.N.
Taroni, C. Tao, Y. Zhang, H. Liu*

TRANSLATIONAL INFORMATICS OF POPULATION HEALTH: HOW LARGE
BIOMOLECULAR AND CLINICAL DATASETS UNITE 455
Y.A. Lussier, A.J. Butte, H. Li, R. Chen, J.H. Moore

Author Index