Lecture Notes in Bioinformatics 10847

Subseries of Lecture Notes in Computer Science

More information about this series at http://www.springer.com/series/5381

Fa Zhang · Zhipeng Cai
Pavel Skums · Shihua Zhang (Eds.)

Bioinformatics Research and Applications

14th International Symposium, ISBRA 2018
Beijing, China, June 8–11, 2018
Proceedings

Springer

Editors
Fa Zhang
Chinese Academy of Sciences
Beijing
China

Zhipeng Cai
Georgia State University
Atlanta, GA
USA

Pavel Skums
Georgia State University
Atlanta, GA
USA

Shihua Zhang
Chinese Academy of Sciences
Beijing
China

ISSN 0302-9743 ISSN 1611-3349 (electronic)
Lecture Notes in Bioinformatics
ISBN 978-3-319-94967-3 ISBN 978-3-319-94968-0 (eBook)
https://doi.org/10.1007/978-3-319-94968-0

Library of Congress Control Number: 2018947451

LNCS Sublibrary: SL8 – Bioinformatics

Printed on acid-free paper

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

On behalf of the Program Committee, we would like to welcome you to the proceedings of the 14th edition of the International Symposium on Bioinformatics Research and Applications (ISBRA 2018), held in Beijing, China, June 8–11, 2018. The symposium provides a forum for the exchange of ideas and results among researchers, developers, and practitioners working on all aspects of bioinformatics and computational biology and their applications.

This year we received 138 submissions in response to the call for extended abstracts. The Program Committee decided to accept 24 of them for full publication in the proceedings and oral presentation at the symposium. We also accepted 30 for oral presentation; a list of these contributions can be found in this front matter. Furthermore, we received ten submissions in response to the call for short abstracts. The technical program also featured two keynote and two invited talks by four distinguished speakers: Prof. Ying Xu from the University of Georgia presented on mining omic data of large numbers of cancer tissue samples; Prof. Xuegong Zhang from Tsinghua University gave a primary view on single-cell bioinformatics; Prof. Xin Gao from King Abdullah University of Science and Technology introduced a graph-based biclustering method for mining phenotype data; Prof. Min Li from Central South University spoke on de novo genome assembly by using statistical characteristics of paired-end reads.

We would like to thank the Program Committee members and the additional reviewers for volunteering their time to review and discuss symposium papers. We would like to extend special thanks to the steering and general chairs of the symposium for their leadership, and to the finance, publicity, workshops, local organization, and publications chairs for their hard work in making ISBRA 2018 a successful event. Last but not least, we would like to thank all authors for presenting their work at the symposium.

April 2018

Fa Zhang
Min Li
Xiaohua Wan
Zhipeng Cai

Organization

Steering Committee

Dan Gusfield	University of California Davis, USA
Ion Mandoiu	University of Connecticut, USA
Yi Pan (Chair)	Georgia State University, USA
Marie-France Sagot	Inria, France
Ying Xu	University of Georgia, USA
Aidong Zhang	State University of New York, USA

General Chairs

Xilin Chen	Institute of Computing Technology, Chinese Academy of Sciences, China
Alexander Zelikovsky	Georgia State University, USA
Shuigeng Zhou	Fudan University, China

Program Chairs

Fa Zhang	Institute of Computing Technology, Chinese Academy of Sciences, China
Zhipeng Cai	Georgia State University, USA
Pavel Skums	Georgia State University, USA
Shihua Zhang	Academy of Mathematics and Systems Science, Chinese Academy of Sciences, China

Publications Chairs

Min Li	Central South University, China
Xiaohua Wan	Institute of Computing Technology, Chinese Academy of Sciences, China
Le Zhang	Southwest University, China
Quan Zhou	Tianjin University, China

Publicity Chairs

Xuan Guo	University of North Texas, USA
Shaoliang Peng	National University of Defense Technology, China

Workshop Chairs

Yudong Liu	Institute of Computing Technology, Chinese Academy of Sciences, China
Fei Ren	Institute of Computing Technology, Chinese Academy of Sciences, China

Webmasters

Sergey Knyazev	Georgia State University, USA
Vyacheslav Tsivina	Georgia State University, USA

Program Committee

Max Alekseyev	George Washington University, USA
Mukul S. Bansal	University of Connecticut, USA
Paola Bonizzoni	Università di Milano-Bicocca, Italy
Zhipeng Cai	Georgia State University, USA
Hongmin Cai	South China University of Technology, China
Doina Caragea	Kansas State University, USA
Xing Chen	National Center for Mathematics and Interdisciplinary Sciences, Chinese Academy of Sciences, China
Xuefeng Cui	Tsinghua University, China
Ovidiu Daescu	University of Texas at Dallas, China
Daming Zhu	Shandong University, China
Fei Deng	University of California, Davis, USA
Lei Deng	Central South University, China
Pufeng Du	Tianjin University, China
Oliver Eulenstein	Iowa State University, USA
Lin Gao	Xidian University, China
Xin Gao	King Abdullah University of Science and Technology, Saudi Arabia
Olga Glebova	Georgia State University, USA
Xuan Guo	University of North Texas, USA
Jiong Guo	Shandong University, China
Zengyou He	Dalian University of Technology, China
Steffen Heber	North Carolina State University, USA
Jinling Huang	East Carolina University, USA
Shihao Ji	Georgia State University, USA
Mingon Kang	Kennesaw State University, USA
Wooyoung Kim	University of Washington Bothell, USA
Danny Krizanc	Wesleyan University, USA

Xiujuan Lei	Shanxi Normal University, China
Min Li	Central South University, China
Shuai-Cheng Li	City University of Hong Kong, SAR China
Yaohang Li	Old Dominion University, USA
Yingshu Li	Georgia State University, Atlanta, USA
Xiaowen Liu	Indiana University-Purdue University Indianapolis, USA
Bingqiang Liu	Shandong University, USA
Ion Mandoiu	University of Connecticut, USA
Fenglou Mao	National Institute of Health, USA
Andrei Paun	University of Bucharest, Romania
Nadia Pisanti	Università di Pisa, Italy
Russell Schwartz	Carnegie Mellon University, USA
Joao Setubal	University of Sao Paulo, Brazil
Xinghua Shi	University of North Carolina at Charlotte, USA
Yi Shi	Shanghai Jiao Tong University, China
Dong Si	University of Washington, USA
Pavel Skums	Georgia State University, USA
Ileana Streinu	Smith College, Northampton, USA
Emily Su	Taipei Medical University, Taiwan
Wing-Kin Sung	National University of Singapore, Singapore
Sing-Hoi Sze	Texas A&M University, USA
Weitian Tong	Georgia Southern University, USA
Gabriel Valiente	Technical University of Catalonia, Spain
Xiaohua Wan	Institute of Computing Technology, Chinese Academy of Sciences, China
Jianxin Wang	Central South University, China
Guohua Wang	Harbin Institute of Technology, China
Li-San Wang	University of Pennsylvania, USA
Seth Weinberg	Virginia Commonwealth University, USA
Fangxiang Wu	University of Saskatchewan, Canada
Yufeng Wu	University of Connecticut, USA
Minzhu Xie	Hunan Normal University, China
Can Yang	Hong Kong Baptist University, SAR China
Kai Ye	Xi'an Jiaotong University, China
Guoxian Yu	Southwest University, China
Alex Zelikovsky	Georgia State University, USA
Fa Zhang	Institute of Computing Technology, Chinese Academy of Sciences, China
Shihua Zhang	Academy of Mathematics and Systems Science, Chinese Academy of Sciences, China
Chi Zhang	Indiana University, USA

Le Zhang	Southwest University, China
Yanqing Zhang	Georgia State University, USA
Jialing Zhang	Yale University, USA
Xing-Ming Zhao	Tongji University, China
Chunhou Zheng	Anhui University, China
Shuigeng Zhou	Fudan University, China
Leming Zhou	University of Pittsburgh, USA
Quan Zou	Tianjin University, China

Abstracts of Invited Talks

A Primary View on Single-Cell Bioinformatics

Xuegong Zhang

School of Life Sciences and School of Medicine, Tsinghua University, Beijing 100084, China
zhangxg@tsinghua.edu.cn

Abstract. Cells are not created equal. The Human Cell Atlas (HCA) project aims to build the atlas of all human cell types and cell states with their molecular signatures. Single-cell sequencing especially single-cell RNA-sequencing (scRNA-seq) is the key technology for obtaining the molecular signatures of a large amount of single cells at the whole transcriptome scale. It is a fundamental step toward the complete understanding of the human body, a super complex system composed of tens of trillions of cells that are all developed from a single cell. This opens the new broad field of single-cell biology. Single-cell biology converts each cell to a mathematical vector in the high-dimensional spaces of the expression of all genes and other molecular features. Therefore, single-cell bioinformatics, or the computational analyses of single-cell data, become the key component of all single-cell biology studies. This talk will give an overview of some key bioinformatics tasks in single-cell bioinformatics, and present examples of our on-going work on new methods for differential expression analysis and dimension reduction.

Searching for Roots of Cancer Development through Mining Large Scale Cancer Tissue Data and Modeling the Chemistry of Cellular Base-Acid Homeostasis

Ying Xu

Department of Biochemistry and Molecular Biology, University of Georgia, Athens, GA, USA
xyn@uga.edu

Abstract. Over one million research articles have been published about cancer, but yet our understanding about cancer is undeniably little. We are yet to understand some of the most basic questions such as: (1) why some cancers such as pancreatic or liver cancers are so deadly while other cancers such as basal cell carcinoma are rarely life-threatening? or (2) why some cancers are highly drug resistant while other cancers are not? In this talk I will present some of our recent discoveries made through mining omic data of large numbers of cancer tissue samples. Our analyses strongly suggest that all cancer tissue cells have high levels of Fenton reactions, due to increased iron accumulation and H2O2 concentration at the disease sites, both being the result of persistent immune responses. A key consequence of the reaction is: it continuously produces OH-, to which the affected cells respond fiercely to maintain the pH homeostasis as changes in the intracellular pH would have profound impacts to the viability of the cells. We will demonstrate that cancer cells immobilize a wide range of metabolic activities through metabolic reprogramming, to keep the intracellular pH stable, including inhibition of the urea cycle, nucleotide synthesis, glycolytic ATP generation (Warburg effect) and even selection of mutations in specific amino acids. Some of the long-standing open questions can be answered naturally using our new model.

Gracob: A Graph-Based Constant-Column Biclustering Method for Mining Growth Phenotype Data

Xin Gao

Computational Bioscience Research Center, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia
xin.gao@kaust.edu.sa

Abstract. Growth phenotype profiling of genome-wide gene-deletion strains over stress conditions can offer a clear picture that the essentiality of genes depends on environmental conditions. Systematically identifying groups of genes from such high-throughput data that share similar patterns of conditional essentiality and dispensability under various environmental conditions can elucidate how genetic interactions of the growth phenotype are regulated in response to the environment. In this talk, I will first demonstrate that detecting such "co-fit" gene groups can be cast as a less well-studied problem in biclustering, i.e., constant-column biclustering. Despite significant advances in biclustering techniques, very few were designed for mining in growth phenotype data. I will then propose Gracob, a novel, efficient graph-based method that casts and solves the constant-column biclustering problem as a maximal clique finding problem in a multipartite graph. We compared Gracob with a large collection of widely used biclustering methods that cover different types of algorithms designed to detect different types of biclusters. Gracob showed superior performance on finding co-fit genes over all the existing methods on both a variety of synthetic data sets with a wide range of settings, and three real growth phenotype data sets for E. coli, proteobacteria, and yeast.

De novo Genome Assembly by Using Statistical Characteristics of Paired-end Reads

Min Li

School of Information Science and Engineering, Central South University, Changsha, China
limin@mail.csu.edu.cn

Abstract. DNA sequence is the carrier of genetic information, which guides the development of biological and functions of life. De novo genome assembly is aimed at acquiring a complete and accurate genome sequence, so it has become one of the fundamental issues in genome research for understanding the organization and process of life activities. However, de novo genome assembly still faces the challenges of repetitive regions in genome, sequencing errors, and uneven sequencing depth. In this talk, I will present our recent work as follows: (1) a sequence assembler based on the distributions of insert size and read, called EPGA. Through assessing the variation of the distribution of insert size, EPGA can solve problems introduced by some complex repetitive regions. And an improved assembler EPGA2 adopts error corrections and memory-efficient DSK to count k-mers; (2) a scaffolding method based on iterative strategy and linear programming to detect spurious edges, called BOSS. And scaffolding algorithm SCOP, which is the first method to classify the contigs and utilize the vertices and edges to optimize the scaffold graph; (3) a gap filling method called GapReduce, which aligns the paired-end reads to the scaffolds. For each gap, GapReduce determines two read sets, and then constructs De Bruijn graphs. GapReduce extracts paths from De Bruijn graphs to cover the gaps by using the characteristics of insert size and k-mer frequencies based on the partitioned read sets. Finally, the future development and challenges of de novo genome assembly will also be discussed.

List of Oral Presentations not Included in this Volume

Drug Repositioning based on Individual Bi-random Walks on a Heterogenous Network
Yuehui Wang, Maozu Guo, Yazhou Ren, Lianyin Jia and Guoxian Yu

Detecting Differential Consistency Network Modules
Jianwei Lu, Yao Lu, Yusheng Ding, Qingyang Xiao, Linqing Liu, Qingpo Cai, Yunchuan Kong, Yun Bai and Tianwei Yu

Joint SNMF Method for Finding Modules of Multiple Brain Networks
Lingkai Tang, Yulian Ding, Jie Zhang and Fang-Xiang Wu

Identifying Driver Genes Involving Gene Dyregulated Expression, Tissue-Specic Expression and Gene-Gene Network
Junrong Song, Feng Wang, Wei Peng and Jianxin Wang

Region-Based Interaction Detection in Genome-Wide Case-Control Studies
Sen Zhang, Wei Jiang, Ronald Cw Ma and Weichuan Yu

HiSSI: High-order SNP-SNP Interactions Detection based on Efficient Significant Pattern and Differential Evolution
Xia Cao, Jie Liu, Maozu Guo and Jun Wang

Detecting Horizontal Gene Transfer: A Probabilistic Approach
Gur Sevillya, Orit Adato and Sagi Snir

Predicting Comorbid Diseases with Geometric Embeddings of Human Interactome
Pakeeza Akram and Li Liao

A Novel Network Based Approach for Predicting Survivability of Breast Cancer Patients
Sheikh Jubair, Luis Rueda and Alioune Ngom

Directional Association Test Reveals High-Quality Putative Cancer Driver Biomarkers Including Noncoding RNAs
Hua Zhong and Mingzhou Song

Mining Information of Co-expression Network based on TGCA Data
Mi-Xiao Hou, Ying-Lian Gao, Jin-Xing Liu, Jun-Liang Shang, Rong Zhu and Sha-Sha Yuan

Identifying MicroRNA-Gene Networks Specific to Pathologic Stages in Colon Cancer
Benika Hall and Xinghua Shi

Cancer Driver Genes Discovery by Integrating Expression and Mutation Data
Ying Hui, Pi-Jing Wei, Junfeng Xia and Chun-Hou Zheng

DDIGIP: Predicting drug-drug interactions based on Gaussian Interaction profile kernels
Cheng Yan, Jianxin Wang, Yi Pan and Fang-Xiang Wu

A Novel Multi-Scale Local Sequence and Structural Representation for Identifying Protein-Protein Interfaces
Fei Guo and Jijun Tang

DEEPSEN: A New Method for Super-Enhancer Prediction based on Convolutional Neural Network
Hongda Bu, Jiaqi Hao, Yanglan Gan, Jihong Guan and Shuigeng Zhou

Data-driven Approach for Understanding the Mild Cognitive Impairment
Bingchen Yu, Meng Han, Liyuan Liu, Yan Huang, Yi Liang and Liquan Bai

InvBFM: Finding Genomic Inversions from High-throughput Sequence Data based on Feature *MiningZhongjia Wu, Yufeng Wu and Jingyang Gao*

Deep Feature Selection with Application to a Neisseria gonorrhoeae Antimicrobial Resistance Study
Jinhong Shi, Yan Yan, Matthew Links, Longhai Li, Michael Horsch and Anthony Kusalik

Jaccard/Tanimoto similarity test and statistical significance estimation methods to evaluate species co-occurrences
Neo Christopher Chung, B lazej Miasojedow, Micha l Startek and Anna Gambin

A Robustness Metric for Biological Data Clustering Algorithms
Yuping Lu, Charles A. Phillips and Michael A. Langston

CSA: A Web Service for the Complete Process of ChIP-Seq Analysis
Min Li, Li Tang, Fang-Xiang Wu, Yi Pan and Jianxin Wang

OffScan: A Universal and Fast CRISPR Off-Target Sites Detection Tool
Yingbo Cui, Minxia Cheng, Jiaming Xu, Xiangke Liao and Shaoliang Peng

GPRED-GC: A Gene PREDiction Model Accounting for 5'–3' GC Gradient
Prapaporn Techa-Angkoon, Kevin Childs and Yanni Sun

Detecting Diagnostic Biomarkers of Alzheimer's Disease by Integrating Gene Expression Data in Six Brain Regions
Lihua Wang and Zhi-Ping Liu

Revealing the Tipping Points in Infant Brain Development for Primates by High Throughput Data
Hui Tang, Ying Tang, Tao Zeng and Luonan Chen

A Framework using Topological Pathways for Deeper Analysis of Transcriptome Data
Yue Zhao, Stephanie Piekos, Tham H. Hoang and Dong-Guk Shin

Detecting Circular RNA from High-throughput Sequence Data with de Bruijn Graph
Xin Li and Yufeng Wu

Identification of Methylation States of DNA Regions for Illumina Methylation BeadChip
Ximei Luo, Yuming Zhao, Fang Wang, and Guohua Wang

Contents

Analysis and Visualization Tools

RNA-Seq Data Analysis