Lecture Notes in Bioinformatics 11488

Subseries of Lecture Notes in Computer Science

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

Ian Holmes · Carlos Martín-Vide ·
Miguel A. Vega-Rodríguez (Eds.)

Algorithms for Computational Biology

6th International Conference, AlCoB 2019 Berkeley, CA, USA, May 28–30, 2019 Proceedings

Springer

Editors
Ian Holmes
University of California, Berkeley
Berkeley, CA, USA

Carlos Martín-Vide
Rovira i Virgili University
Tarragona, Spain

Miguel A. Vega-Rodríguez
University of Extremadura
Cáceres, Spain

ISSN 0302-9743 ISSN 1611-3349 (electronic)
Lecture Notes in Bioinformatics
ISBN 978-3-030-18173-4 ISBN 978-3-030-18174-1 (eBook)
https://doi.org/10.1007/978-3-030-18174-1

LNCS Sublibrary: SL8 – Bioinformatics

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

Preface

These proceedings contain the papers that were presented at the 6th International Conference on Algorithms for Computational Biology (AlCoB 2019), held in Berkeley, California, USA, during May 28–30, 2019.

The scope of AlCoB includes topics of either theoretical or applied interest, namely:

- Sequence analysis
- Sequence alignment
- Sequence assembly
- Genome rearrangement
- Regulatory motif finding
- Phylogeny reconstruction
- Phylogeny comparison
- Structure prediction
- Compressive genomics
- Proteomics: molecular pathways, interaction networks, mass spectrometry analysis
- Transcriptomics: splicing variants, isoform inference and quantification, differential analysis
- Next-generation sequencing: population genomics, metagenomics, metatranscriptomics, epigenomics
- Genome CD architecture
- Microbiome analysis
- Cancer computational biology
- Systems biology

AlCoB 2019 received 30 submissions. Most papers were reviewed by three Program Committee members. There were also a few external reviewers consulted. After a thorough and vivid discussion phase, the committee decided to accept 15 papers (which represents an acceptance rate of about 50%). The conference program included five invited talks and some poster presentations of work in progress.

The excellent facilities provided by the EasyChair conference management system allowed us to deal with the submissions successfully and handle the preparation of these proceedings in time.

We would like to thank all invited speakers and authors for their contributions, the Program Committee and the external reviewers for their cooperation, and Springer for its very professional publishing work.

March 2019

Ian Holmes
Carlos Martín-Vide
Miguel A. Vega-Rodríguez

Organization

AlCoB 2019 was organized by the University of California, Berkeley, USA, and the Institute for Research Development, Training and Advice (IRDTA), Brussels/London, Belgium/UK.

Program Committee

Can Alkan	Bilkent University, Turkey
Stephen Altschul	National Center for Biotechnology Information, USA
Philipp Bucher	Swiss Institute for Experimental Cancer Research, Switzerland
Ken Chen	MD Anderson Cancer Center, USA
Keith A. Crandall	George Washington University, USA
Colin Dewey	University of Wisconsin, Madison, USA
Eytan Domany	Weizmann Institute of Science, Israel
Robert Edgar	Independent, USA
Dmitrij Frishman	Technical University of Munich, Germany
Susumu Goto	Research Organization of Information and Systems, Japan
Desmond Higgins	University College Dublin, Ireland
Karsten Hokamp	Trinity College Dublin, Ireland
Ian Holmes	University of California, Berkeley, USA
Fereydoun Hormozdiari	University of California, Davis, USA
Daniel Huson	University of Tübingen, Germany
Martijn Huynen	Radboud University Medical Centre, The Netherlands
Peter Karp	SRI International, USA
Kazutaka Katoh	Osaka University, Japan
Anders Krogh	University of Copenhagen, Denmark
Doron Lancet	Weizmann Institute of Science, Israel
Alla Lapidus	Saint Petersburg State University, Russia
Ming Li	University of Waterloo, Canada
Gerard Manning	Genentech, USA
Carlos Martín-Vide (Chair)	Rovira i Virgili University, Spain
David H. Mathews	University of Rochester Medical Center, USA
Aaron McKenna	University of Washington, USA
Jason Rafe Miller	Shepherd University, USA
Aleksandar Milosavljevic	Baylor College of Medicine, USA
Yasukazu Nakamura	National Institute of Genetics, Japan
Zemin Ning	Wellcome Trust Sanger Institute, UK
William Stafford Noble	University of Washington, USA
Sandra Orchard	European Bioinformatics Institute, UK

William Pearson	University of Virginia, USA
Matteo Pellegrini	University of California, Los Angeles, USA
Mihaela Pertea	Johns Hopkins University, USA
Steve Rozen	Duke-NUS Medical School, Singapore
David Sankoff	University of Ottawa, Canada
Russell Schwartz	Carnegie Mellon University, USA
Wing-Kin Sung	National University of Singapore, Singapore
Alfonso Valencia	Barcelona Supercomputing Centre, Spain
Arndt von Haeseler	Center for Integrative Bioinformatics Vienna, Austria
Kai Wang	Children's Hospital of Philadelphia, USA

Additional Reviewers

Wout Bittremieux	Shaoheng Liang
Viraj Deshpande	Vakul Mohanty
Petko Fiziev	Jacob Schreiber
Markus Fleischauer	Yifei Shen
Songling Li	Ritambhara Singh

Organizing Committee

Ian Holmes (Co-chair)	University of California, Berkeley, USA
Sara Morales	IRDTA, Brussels, Belgium
Manuel Parra-Royón	University of Granada, Granada, Spain
David Silva (Co-chair)	IRDTA, London, UK
Miguel A. Vega-Rodríguez	University of Extremadura, Cáceres, Spain

Abstracts of Invited Talks

Exploring Phenotypic Heterogeneity Across Tissues and Conditions with Network-Based Approaches

Teresa M. Przytycka

National Center of Biotechnology Information, National Library of Medicine, NIH, Bethesda, MD 20894, USA

Phenotypic heterogeneity is assumed arise as the result of a combination of genetic, epigenetic, and environmental factors and the stochastic nature of biochemical processes, such as gene expression events, during the development. In the last few years, my group has studied phenotypic heterogeneity in two different contexts: in the context of cancer [2–5, 7, 8] and in the context of the model organism – *Drosophila melanogaster* [10, 11].

Functional interaction networks, that is networks whose edges represent functional relationships between genes, provide an important context for studies of organismal phenotypes. A network-centric view of genotype-phenotype relation proposes that perturbing functionally related genes is likely to lead to similar phenotypes. Indeed, network-centric approaches have proven to be helpful for finding genotypic causes of diseases, classifying disease into subtypes, and identifying drug targets [1, 5]. To support such pathway-centric perspective, algorithms that leverage biological networks to advance the understanding of phenotypic heterogeneity are necessary.

In our earlier study, building on the set cover approach, we have developed network-based approaches to identify pathways dysregualted in cancer [4, 7]. In contrast, focusing on uncovering pathways mediating the relation between somatic mutations and dysregulated gene expression modules, we utilised a network flow approach [6, 8]. Complementing these studies, our recently developed method, BeWith, utilises Integer Linear Programming and an integrated analysis of mutual exclusivity, co-occurrence and functional interaction networks to uncover the relationships between mutated gene modules. As expected such mutated gene modules often underline specific cancer sub-types [3].

In addition to studies on dysregulated pathways in cancer and cancer subtypes, network based approaches can also shed light on other phenotypes such as drug response. Towards this end we have recently developed NETPHLIX - an algorithm to identify mutated subnetworks that are associated with a continuous phenotype. Subsequently, we utilised NETPHLIX to identify mutated gene networks that are associated with response to drugs. Another recently emerged phenotype in cancer studies is the presence and strength of the so-called mutational signatures. Mutational signatures are indicative of mutagenic processes that have been active in the given patient. Such processes are often triggered by genetic causes such as a dysfunctional DNA repair pathway. Understanding the mechanism behind the emergence of a particular

mutational signature is challenging since increased mutagenic activities leads to an increased amount of passenger mutations making it difficult to untangle the cause from the effect. Using NETPHLIX, we were able to identify mutated sub-networks associated with several mutational signatures in breast cancer [9].

In contrast to functional interaction networks, gene regulatory networks (GRN) summarise regulatory relationships between transcription factors (TF) and the gens that they regulate. GRN regulate maintenance of cell type specific states, response to stress, and other cell functions. Thus phenotypic differences can be potentially explained by differences in gene regulation. However methods to infer GRN are typically context-agnostic. To address this challenge, we have recently introduced a novel computation method NetREX that given a context-agnostic network as a prior and context specific expression data (for example data for a healthy and a disease tissue), constructs context-specific GRNs by rewiring the prior network [11]. Comparative analysis of such networks can provide yet another window to study phenotypic differences.

We conclude that network based approaches, supported by a variety of algorithmic approaches can provide important stepping stone towards understating phenotypic heterogeneity.

Acknowledgements. I would like to acknowledge all the collaborators of the work discussed in this talk. Particular thanks to the current and former members of my group: Yoo-Ah Kim, Yijie Wang, Damian Wojtowicz, Phoung Dao, Jan Hoinka, Dang-Yon Cho, Raheleh Salari our visiting group member Rebecca Sarto-Basso, and our many collaborators including Roded Sharan, Fabio Vandin, Brian Oliver, Dorit S Hochbaum, Stefan Wuchty, Hang Noh Lee, Max Leiserson and all the other collaborators listed in the bibliography of this talk. The research in Przytycka's group is supported by the Intramural Research Programs of the National Library of Medicine at National Institutes of Health, USA.

References

1. Cho, D.Y., Kim, Y.A., Przytycka, T.: Network biology approach to complex diseases. PLoS Comput. Biol. **8**(12) (2012). https://doi.org/10.1371/journal.pcbi.1002820
2. Cho, D.Y., Przytycka, T.M.: Dissecting cancer heterogeneity with a probabilistic genotype-phenotype model. In: Deng, M., Jiang, R., Sun, F., Zhang, X. (eds.) RECOMB 2013. LNCS, vol. 7821, pp. 30–31. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-37195-0_3
3. Dao, P., Kim, Y.A., Wojtowicz, D., Madan, S., Sharan, R., Przytycka, T.: BeWith: a between-within method to discover relationships between cancer modules via integrated analysis of mutual exclusivity, co-occurrence and functional interactions. PLoS Comput. Biol. **13**(10) (2017). https://doi.org/10.1371/journal.pcbi.1005695
4. Kim, Y.A., Cho, D.Y., Dao, P., Przytycka, T.: MEMCover: integrated analysis of mutual exclusivity and functional network reveals dysregulated pathways across multiple cancer types. Bioinformatics **31**(12) (2015). https://doi.org/10.1093/bioinformatics/btv247
5. Kim, Y.A., Cho, D.Y., Przytycka, T.: Understanding genotype-phenotype effects in cancer via network approaches. PLoS Comput. Biol. **12**(3) (2016). https://doi.org/10.1371/journal.pcbi.1004747

6. Kim, Y.A., Przytycki, J., Wuchty, S., Przytycka, T.: Modeling information flow in biological networks. Phys. Biol. **8**(3) (2011). https://doi.org/10.1088/1478-3975/8/3/035012
7. Kim, Y.A., Salari, R., Wuchty, S., Przytycka, T.: Module cover - a new approach to genotype-phenotype studies. In: 18th Pacific Symposium on Biocomputing, PSB 2013 (2013)
8. Kim, Y.A., Wuchty, S., Przytycka, T.: Identifying causal genes and dysregulated pathways in complex diseases. PLoS Comput. Biol. **7**(3) (2011). https://doi.org/10.1371/journal.pcbi.1001095
9. Kim, Y.A., et al.: Network-based approaches elucidate differences within APOBEC and clock-like signatures in breast cancer. bioRxiv, p. 568568, March 2019. https://doi.org/10.1101/568568, https://www.biorxiv.org/content/10.1101/568568v1?rss=1
10. Lee, H., et al.: Dosage-dependent expression variation suppressed on the Drosophila male X chromosome. G3: Genes Genomes Genet. **8**(2) (2018). https://doi.org/10.1534/g3.117.300400
11. Wang, Y., Cho, D.Y., Lee, H., Fear, J., Oliver, B., Przytycka, T.: Reprogramming of regulatory network using expression uncovers sex-specific gene regulation in Drosophila. Nat. Commun. **9**(1) (2018). https://doi.org/10.1038/s41467-018-06382-z

New Divide-and-Conquer Techniques for Large-Scale Phylogenetic Estimation[1]

Tandy Warnow

Department of Computer Science, University of Illinois at Urbana-Champaign, 201 N. Goodwin Ave, Urbana, IL, 61801, USA
warnow@illinois.edu

Over the last years, the availability of genomic sequence data from thousands of different species has led to hopes that a phylogenetic tree of all life might be achievable. Yet, the most accurate methods for estimating phylogenies are heuristics for NP-hard optimization problems, many of which are too computationally intensive to use on large datasets. Divide-and-conquer approaches have been proposed to address scalability to large datasets that divide the species into subsets, construct trees on subsets, and then merge the trees together. Prior approaches have divided species sets into overlapping subsets and used supertree methods to merge the subset trees, but limitations in supertree methods suggest this kind of divide-and-conquer approach is unlikely to provide scalability to ultra-large datasets. Recently, a new approach has been developed that divides the species dataset into disjoint subsets, computes trees on subsets, and then combines the subset trees using auxiliary information (e.g., a distance matrix). Here, we describe these strategies and their theoretical properties, present open problems, and discuss opportunities for impact in large-scale phylogenetic estimation using these and similar approaches.

Acknowledgements. This work was supported in part by NSF grant CCF-1535977. I also wish to thank Erin Molloy and Thien Le for helpful comments on the manuscript.

[1] Supported by the University of Illinois at Urbana-Champaign.

Contents