Essentials of Bioinformatics, Volume II

Noor Ahmad Shaik • Khalid Rehman Hakeem
Babajan Banaganapalli • Ramu Elango
Editors

Essentials of Bioinformatics, Volume II

In Silico Life Sciences: Medicine

Springer

Editors
Noor Ahmad Shaik
Princess Al-Jawhara Center of Excellence
in Research of Hereditary Disorders
Department of Genetic Medicine
Faculty of Medicine
King Abdulaziz University
Jeddah, Saudi Arabia

Babajan Banaganapalli
Princess Al-Jawhara Center of Excellence
in Research of Hereditary Disorders
Department of Genetic Medicine
Faculty of Medicine
King Abdulaziz University
Jeddah, Saudi Arabia

Khalid Rehman Hakeem
Department of Biological Sciences
King Abdulaziz University
Jeddah, Saudi Arabia

Princess Dr. Najla Bint Saud Al-Saud
Center for Excellence
Research in Biotechnology
King Abdulaziz University
Jeddah, Saudi Arabia

Ramu Elango
Princess Al-Jawhara Center of Excellence
in Research of Hereditary Disorders
Department of Genetic Medicine
Faculty of Medicine
King Abdulaziz University
Jeddah, Saudi Arabia

ISBN 978-3-030-18374-5 ISBN 978-3-030-18375-2 (eBook)
https://doi.org/10.1007/978-3-030-18375-2

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

Dr. Ramu Elango *dedicates the book to his wife, Karpagam; their children, Madhivanan and Ilankeeran; and his brothers, Govindan and Prof Krishnan Ramu and Anni-Vijaya Krishnan*

Dr. Babajan Banaganapalli *dedicates the book to his family, Khader Basha, Shanu, Shanawaz, Shavar, and Shaju, wife, Gowsia; and their children, Rayyan and Reenah*

Dr. Noor Ahmad Shaik *dedicates the book to his parents, Jan Ahmad and Abida Khatoon; siblings (Nazmul, Jeelani, Reshma and Nazia); wife, Dr. Fatima; and children, Maryam, Sakina, and Raheel*

Dr. Khalid Rehman Hakeem *dedicates the book to his beloved family: Papa, Mom, Asma, and Hibah*

for their sacrifices and continued support.

Foreword

The golden era of the biological understanding of the health and disease is now unfolding the nature secrets in an unprecedented level. This revolution is not possible without the contributions of the scientists across the world as well as in many subjects, ranging from biologist; physicists; engineers of the variety of fields like electrical, mechanical, and computer science; as well innovators of new ideas. This free flow of ideas from people with different skills resulted in bringing the new technologies and their implementation in a way which was never in doubt changed the way we look at the life around as well as inside us from the environment and the interaction with genomes resulting in adaption and better management plans of our lives with improved health.

Biomedical scientists were the major beneficiaries of such advances of the variety of fields mentioned above. Almost 50 years separated the discovery of DNA structure to the Human Genome Project achievement. Thousands of scientists developed new methods and technologies from sequencing and computers with power to deal with complex data generation to analysis. The proverb "Necessity is the mother of invention" explains the development of bioinformatics aptly. For example, with the generation of sequences came the first step of theoretical scientists, mathematicians and statisticians put the first seed for BLAST, to compare different sequences now to deal with high-throughput data from a spectrum of "-omic" technologies like genomics, proteomics, and metabolomics. Comparing fewer than 1000 bases in early years to millions and billions of bases and data points in biology in a short time with a variety of tools resulted in the rapid development in diagnosis of many genetic defects in rare diseases to identification of hundreds and thousands of risk markers for the complex diseases plaguing the human race at an alarming rate. Now, thanks to the bioinformatics tool, biologist with limited or no knowledge of computer programs can analyze the complex data from a variety of high-throughput "-omic" fields to search for the answer to their scientific queries.

This book series is trying to target the graduate students and young researchers who are keen in understanding and contemplating their future career in high-throughput biological fields of their choice. The chapters give the flavor of the various fields from genetic diagnosis, the dissection of complex diseases to

application, and the collaborative efforts of bioinformatics scientists with geneticists, statisticians, biochemists, and engineers to deliver the new understanding of the human biology. The first volume showed a variety of tools available in bioinformatics field to address a variety of queries with different sets of biological data. The current second volume gives a glimpse of the success of such technologies and bioinformatics tools in many fields, changing the disease diagnostics and novel drug identification to better patient management with better drugs and exploring the revolutionary stem cell science to treat patients of devastating diseases. The editors were bold enough to take the task of assembling a group of senior scientists and young people, who understand the need and difficulties of young researchers and graduate students, in unravelling the myth that advanced biological research is unreachable to them in a simple format. I congratulate the senior and experienced authors of various chapters and editors for providing an excellent overview, highlighting the impact of bioinformatics and "-omic" technologies across many fields to improve the human welfare.

I strongly recommend this volume series to the young students and budding researchers wishing to enter this exciting era of biomedical revolutionary research. I am confident this series of volumes will provide the confidence to science students in different corners of the world, especially from the developing world with limited resources, to dream up the careers in this field to make an impact on the world.

Prof. Kaipa Prabhakar Rao
Department of Genetics
Osmania University
Hyderabad, India

Preface

Bioinformatics is growing along with the rapid advances in many different technological and scientific fields. The "big data" science is the result of combined work of ultrahigh-throughput technology development and high-performance computers. Genetics, genomics, proteomics, metabolomics, and metagenomics changed the biology more in the recent past. Next-generation sequencing technology is the result of Human Genome Project with whole-exome and whole-genome sequencing (NGS) possible within 24–56 hours. This revolutionized the genetic diagnosis of rare diseases around the world. Almost every country has the scientists equipped with the NGS data analysis skills for diagnostic purposes. Bioinformatics tools, especially in the public domain, make this technology for research and application in diagnosis a reality in every corner of the world. Many nations realized the potential of the national biobank and their potential contribution to the economy by reducing the healthcare burden enormously through the prevention of disease and/or better management of patients through novel drug discovery to personalized medicine.

This volume, like the first volume, is targeting the young researchers to make them aware of the recent developments in a variety of fields where bioinformatics along with the other multi-omics technologies changed the scientific world, making a large impact. It also focusses on the key development in key multi-omics technologies output and their impact in many aspects of biomedical fields. Human genome sequencing project witnessed a heightened activity of bioinformatics scientists and tools. Hundreds and thousands of the easy tools were developed for a variety of applications. It is not possible to discuss the examples for any single group of bioinformatics tools. Hand in hand with the first volume, this will help the young scientists and graduate students realize the role of bioinformatics play in the development of many applied biomedical advances toward better healthcare for all.

The chapters are organized in a way to highlight a particular "-omic" technology and its role in changing the biomedical scientific area. For example, the microarray and NGS technology, combined with the bioinformatics tools, made the genetic diagnosis rapid and accurate, even for rare diseases in any corner of the world with very little blood within days. Unknown diseases reveal novel hidden mutations,

helping the scientists learn more about the disease biology to address the biological understanding in finer detail. Likewise, drug discovery and personalized medicine had the bioinformatics stamped its impact along with the technologies. We, as scientists, attempted to highlight the success of various biomedical fields in this volume to support the role of collaborative nature of modern science among the multidisciplinary scientists. It is the celebrations of the collaborative scientists ranging from physical to applied medical and clinical scientists with bioinformatics groups, directly or indirectly, through many software tools or specialized databases with hidden tools to provide the accurate answer to their queries. Hopefully, the young scientists will realize the importance of this type of multidisciplinary collaboration and gain success in their professional careers.

We sincerely thank the management, faculty members, staff, and students at Princess Al-Jawhara Center of Excellence in Research of Hereditary Disorders (PACER-HD), Department of Genetic Medicine, Faculty of Medicine, and Department of Biology, Faculty of Science, at King Abdulaziz University (KAU) for supporting our effort in bringing this book series a reality. Our special thanks go to Prof. Jumana Y. Al-Aama, director of PACER-HD, KAU, for letting us realize the importance of bioinformatics in clinical practice, for encouraging excellent scientific discussions and raising critical questions as clinicians, and for supporting our work throughout this long process. We would also like to thank the chairman of the Department of Biological Sciences, Prof. Khalid M. AlGhamdi, and the head of Plant Sciences Section, Dr. Hesham F. Alharby, for providing us the valuable suggestions and encouragement to complete this task. We also acknowledge the authors of all chapters who spared their precious time in bringing this book out with valuable contributions. Last but not the least, we would like to acknowledge Springer Nature publishers, especially Mr. Rahul Sharma, for their patience and regular communication with us to move the project forward.

Jeddah, Saudi Arabia — Noor Ahmad Shaik
Jeddah, Saudi Arabia — Khalid Rehman Hakeem
Jeddah, Saudi Arabia — Babajan Banaganapalli
Jeddah, Saudi Arabia — Ramu Elango

Contents

About the Editors

Noor Ahmad Shaik is an academician, researcher, and technologist working in the field human molecular genetics. Over the last 15 years, he has been working with different research groups whose fundamental focus is to offer genetic disease diagnostics, management, and therapy. Currently, he is working to improve the current mutation prediction methods through integrative computational algorithms, so that clinicians and scientists can better understand the functional relevance of genetic mutations to disease, and is rendering his editorial services to world-renowned journals like the *Frontiers in Pediatrics* and *Frontiers in Genetic Disorders*. He is interested in discovering the novel causal genes/biomarkers for rare hereditary disorders and also in understanding the effect of mutations on structure and function of causal proteins of human diseases. He has already published 42 research publications in reputed international journals of human genetics and bioinformatics and has been a recipient of several research grants from national and international funding agencies.

Khalid Rehman Hakeem is Professor at King Abdulaziz University, Jeddah, Saudi Arabia. He has completed his PhD (Botany) from Jamia Hamdard, New Delhi, India, in 2011 and has worked as postdoctorate fellow in 2012 and fellow researcher (associate prof.) from 2013 to 2016 at the Universiti Putra Malaysia, Selangor, Malaysia. His specialty includes Plant Ecophysiology, Biotechnology and Molecular Biology, Plant-Microbe-Soil Interactions, and Environmental Sciences. So far, he has edited and authored more than 25 books with Springer International, Academic Press (Elsevier), CRC Press, etc. and has also, to his credit, published more than 120 research publications in peer-reviewed international journals, including 42 book chapters in edited volumes with the international publishers.

Babajan Banaganapalli works as bioinformatics research faculty at King Abdulaziz University, where he initiated and successfully run the interdisciplinary Bioinformatics program from 2014 till to date. He has more than 12 years of research experience in bioinformatics; has published more than 40 journal articles, conference papers, and book chapters; and has also served in numerous conference program committees, organized several bioinformatics workshops and training programs, and acted as editor and reviewer for various international genetics/bioinformatics journals. His research interests spread across genomics, proteomics, and drug discovery for complex diseases. Recently, he was honored as young scientist for his outstanding research in bioinformatics by Venus International Research Foundation, India.

Ramu Elango is a well-experienced molecular geneticist and computational biologist with extensive experience at MIT, Cambridge, USA, and GlaxoSmithKline R&D, UK, where he contributed extensively in many disease areas of interest in identifying novel causal genes and tractable drug targets, after completing his PhD in Human Genetics at All India Institute of Medical Sciences, New Delhi, India. He presently heads the Research and Laboratories at the Princess Al-Jawhara Center of Excellence in Research of Hereditary Disorders, King Abdulaziz University. His research focus is on genetics and genomics of complex and polygenic diseases. His team exploits freely available large-scale genetic and genomic data with bioinformatics tools to identify the risk factors or candidate causal genes for many complex diseases.