

Essentials of Genomics and Bioinformatics

Edited by C. W. Sensen

Essentials of Genomics and Bioinformatics

Edited by C. W. Sensen

 **WILEY-VCH**

Edited by:
Prof. Christoph W. Sensen
University of Calgary
Faculty of Medicine
Department of Biochemistry and
Molecular Biology
3330 Hospital Drive N.W.
Calgary, Alberta
Canada

This book was carefully produced. Nevertheless, authors, editor and publisher do not warrant the information contained therein to be free of errors. Readers are advised to keep in mind that statements, data, illustrations, procedural details or other items may inadvertently be inaccurate.

Library of Congress Card No. applied for

A catalogue record for this book is available from the British Library

Die Deutsche Bibliothek – Cataloguing-in-Publication Data:

A catalogue record for this book is available from Die Deutsche Bibliothek

ISBN 3-527-30541-6

© WILEY-VCH Verlag GmbH, Weinheim, 2002

Printed on acid-free paper.

All rights reserved (including those of translation in other languages). No part of this book may be reproduced in any form – by photoprinting, microfilm, or any other means – nor transmitted or translated into a machine language without written permission from the publishers. Registered names, trademarks, etc. used in this book, even when not specifically marked as such, are not to be considered unprotected by law.

Composition and Printing: Zechner Datenservice und Druck, Speyer.

Bookbinding: Grossbuchbinderei J. Schäffer GmbH + Co. KG, Grünstadt.

Printed in the Federal Republic of Germany

Preface

Genomics has become the new paradigm in biological and medical research. Initially being defined as the large-scale characterization of genomes (mainly through mapping and DNA sequencing), the field has become increasingly diverse, now covering the characterization of genomes on the sequence level, as well as the large-scale study of gene expression and the protein complement of organisms. Due to the rapid development of the field, it is quite difficult even for the specialist to stay informed about the wealth of new approaches that is being developed every year. This applies even more to the interested general public, thus this book tries to provide comprehensive introductory level information about the methods used in Genomics research, the model organisms studied, the bioinformatics approaches, which are used to analyze genomic data, and the ethical implications of genome research.

The first section of the book is dedicated to model organisms, including humans and the study of their genomes. Several chapters in this section deal with the characterization and analysis of the human genome, and the potential of Genomics for the cure of diseases. The next two sections of the book deal with DNA and protein technologies, including chapters on genomic mapping, DNA sequencing, DNA high-density arrays and the analysis of the proteome through Mass Spectrometry and Capillary Electrophoresis. The following section

contains six chapters on the bioinformatics aspects of Genomics. While many books now deal with bioinformatics in general, these chapters are different as they exclusively deal with the analysis of complete genomes, a topic rarely covered elsewhere. The book concludes with a chapter about the ethical implications of genome research and an outlook to the future of Genomics.

This book started out as part of the Second Edition of *Biotechnology*. The success of Volume 5b *Genomics and Bioinformatics* encouraged the series editors, the publisher and myself to pursue a “spin off” from the series, which resulted in the current revised concise softcover edition. I would like to thank H.-J. Rehm, G. Reed, A. Pühler, P. Stadler and WILEY-VCH for the permission to let *Biotechnology* Volume 5b be used as the basis for this book. Special thanks go to Karin Dembowski, who has now patiently guided the creation of this book in an extremely competent manner for over two years. Without her constant encouragement and support, it would have been impossible to complete this book. Finally, we are grateful for the suggestions provided by Stephanie E. Minnema (The University of Calgary, Faculty of Medicine) after critically reading *Biotechnology* Volume 5b *Genomics and Bioinformatics*.

Calgary, January 2002

C. W. Sensen

Contents

Introduction 1

C. W. Sensen

Application Domains

1 Genome Projects of Model Organisms 5

*A. Pühler, D. Jording, J. Kalinowski,
D. Buttgerit, R. Renkawitz-Pohl,
L. Altschmied, A. E. Danchin,
H. Feldmann, H.-P. Klenk, M. Kröger*

2 The Human Genome Project 41

L. C. Tsui, S. W. Scherer

3 Genetic Diseases 61

G. Dellaire

4 Pharmaceutical Bioinformatics and Drug Discovery 105

C. S. Ramanathan, D. B. Davison

5 Agri-Food and Genomics 127

*I. A. P. Parkin, S. J. Robinson, A. G. Sharpe,
K. Rozwadowski, D. Hegedus, D. J. Lydiate*

DNA Technologies

6 Genomic Mapping and Positional Cloning, with Emphasis on Plant Science 145

*A. Vanavichit, S. Tragoonrun,
T. Toojinda*

7 Sequencing Technology 165

*L. R. Middendorf, P. G. Humphrey,
N. Narayanan, S. C. Roemer*

8 A DNA Microarray Fabrication Strategy for Research Laboratories 209

*D. C. Tessier, F. Benoit, T. Rigby, H. Hogues,
V. Van het Hoog, D. Y. Thomas,
R. Brousseau*

Protein Technologies

9 Two-Dimensional Gel Electrophoresis and Mass Spectrometry for Proteomic Studies: State-of-the-Art 223

D. Figeys

10 Proteome Analysis by Capillary Electro- phoresis 253

*N. J. Dovichi, Shen Hu, D. Michels,
Z. Zhang, S. N. Krylov*

Bioinformatics

11 Using the Molecular Biology Data 263

*E. M. Zdobnov, R. Lopez, R. Apweiler,
T. Etzold*

12 Tools for DNA Technologies 285

P. Rice

13 Tools for Protein Technologies 309

D. S. Wishart

14 Structure Information 329

M. Cygler, A. Matte, J. D. Schrag

VIII Contents

- 15 Genomic Data Representation through
Images – MAGPIE as an Example 345
P. Gordon, T. Gaasterland, C. W. Sensen
- 16 Interaction Databases 365
G. D. Bader, C. W. V. Hogue

Ethical, Legal and Social Issues

- 17 Ethical Aspects of Genome Research and
Banking 381
B. M. Knoppers
- 18 Genomics – Five Years from Now 397
C. W. Sensen

Index 407

Color Figures *see*

[http://www.wiley-vch.de/books/info/
3-527-30541-6](http://www.wiley-vch.de/books/info/3-527-30541-6)

Contributors

Dr. Lothar Altschmied
AG Expressionskartierung
Institut für Pflanzengenetik und
Kulturpflanzenforschung (IPK)
Corrensstraße 3
D-06466 Gatersleben
Germany
Chapter 1

Dr. Rolf Apweiler
EMBL Outstation
The European Bioinformatics Institute
Hinxton Hall, Hinxton
Cambridge, CB10 1SD
United Kingdom
Chapter 11

Gary D. Bader
Samuel Lunenfeld Research Institute
Mount Sinai Hospital
600 University Avenue
Toronto, Ontario, M5G 1X5
Canada
Chapter 16

Dr. Roland Brousseau
National Research Council of Canada
6100 Royalmount Avenue
Montreal, Quebec, H4P 2R2
Canada
Chapter 8

Dr. Detlev Buttgereit
Zell- und Entwicklungsbiologie
FB Biologie
Universität Marburg
Karl-von-Frisch-Straße
D-35032 Marburg
Germany
Chapter 1

Dr. Mirosław Cygler
Biotechnology Research Institute
6100 Royalmont Avenue
Montreal, Quebec, H4P 2R2
Canada
Chapter 14

Prof. Dr. Antoine Danchin
Hong Kong University,
Pasteur Research Centre
Dexter HC Man Building
8, Sassoon Road, Pokfulam
Hong Kong
and
Genetics of Bacterial Genomes,
CNRS URA 2171
Institut Pasteur
28, rue du Docteur Roux
75724 Paris Cedex 15
France
Chapter 1

Dr. Daniel B. Davison
Director of Bioinformatics
Bristol Myers Squibb
Bioinformatics-853
5 Research Parkway
Wallingford, CT 06492-7660
USA
Chapter 4

Dr. Graham Dellaire
MRC – Human Genetics Unit
Crewe Road
Edinburgh, EH4 2XU
United Kingdom
Chapter 3

Dr. Norman Dovichi
Department of Chemistry
University of Washington
Seattle, WA 98195-1700
USA
Chapter 10

Dr. Thure Etzold
Lion Bioscience
Compass House
80–82 Newmarket Road
Cambridge, CB5 8DZ
United Kingdom
Chapter 11

Prof. Dr. Horst Feldmann
Adolf-Butenandt-Institut
Schillerstraße 44
D-80336 München
Germany
Chapter 1

Dr. Daniel Figeys
MDS Proteomics Inc.
251 Attwell Drive
Toronto, Ontario, M9W 7H4
Canada
Chapter 9

Dr. Terry Gaasterland
The Rockefeller University
1230 York Avenue
New York, NY 10021-6399
USA
Chapter 15

Paul Gordon
National Research Council of Canada
1411 Oxford Street
Halifax, Nova Scotia, B3H 3Z1
Canada
Chapter 15

Dr. Dwayne Hegedus
AAFC, Saskatoon Research Center
Molecular Genetics Section
107 Science Place
Saskatoon, Saskatchewan, S7N 0X2
Canada
Chapter 5

Dr. Christopher Hogue
Samuel Lunenfeld Research Institute
Mount Sinai Hospital
600 University Avenue
Toronto, Ontario, M5G 1X5
Canada
Chapter 16

Dr. Shen Hu
Department of Chemistry
University of Washington
Seattle, WA 98195-1700
USA
Chapter 10

Dr. Patrick G. Humphrey
Advanced Research and Development
Li-Cor, Inc.
4308 Progressive Avenue
P.O. Box 4000
Lincoln, NE 68504
USA
Chapter 7

Dr. Doris Jording
Lehrstuhl für Genetik
Universität Bielefeld
D-33594 Bielefeld
Germany
Chapter 1

Dr. Jörn Kalinowski
Lehrstuhl für Genetik
Universität Bielefeld
D-33594 Bielefeld
Germany
Chapter 1

Dr. Hans-Peter Klenk
EPIDAUROS Biotechnologie AG
D-82347 Bernried
Germany
Chapter 1

Prof. Bartha Maria Knoppers
Université de Montréal
Faculté de Droit
3101 chemin de la Tour
C.P. 6128, succursale A
Montreal, Quebec, H3C 3J7
Canada
Chapter 17

Prof. Dr. Manfred Kröger
Institut für Mikro- und Molekularbiologie
Universität Gießen
Heinrich-Buff-Ring 26–32
D-35392 Gießen
Germany
Chapter 1

Dr. Sergej N. Krylov
Department of Chemistry
York University
Toronto, Ontario, M3J 1P3
Canada
Chapter 10

Dr. Rodrigo Lopez
EMBL Outstation
The European Bioinformatics Institute
Hinxton Hall, Hinxton
Cambridge, CB10 1SD
United Kingdom
Chapter 11

Dr. Derek Lydiate
AAFC, Saskatoon Research Center
Molecular Genetics Section
107 Science Place
Saskatoon, Saskatchewan, S7N 0X2
Canada
Chapter 5

Dr. Alan Matte
Biotechnology Research Institute
6100 Royalmont Avenue
Montreal, Quebec, H4P 2R2
Canada
Chapter 14

Dr. David Michels
Department of Chemistry
University of Washington
Seattle, WA 98195-1700
USA
Chapter 10

Dr. Lyle R. Middendorf
Advanced Research and Development
Li-Cor, Inc.
4308 Progressive Avenue
P.O. Box 4000
Lincoln, NE 68504
USA
Chapter 7

Dr. Narasimhachari Narayanan
Advanced Research and Development
Li-Cor, Inc.
4308 Progressive Avenue
P.O. Box 4000
Lincoln, NE 68504
USA
Chapter 7

Dr. Isobel A. P. Parkin
AAFC, Saskatoon Research Center
Molecular Genetics Section
107 Science Place
Saskatoon, Saskatchewan, S7N 0X2
Canada
Chapter 5

Prof. Dr. Alfred Pühler
Lehrstuhl für Genetik
Universität Bielefeld
D-33594 Bielefeld
Germany
Chapter 1

Dr. Chandra S. Ramanathan
Bristol Myers Squibb
Bioinformatics-853
5 Research Parkway
Wallingford, CT 06492-7660
USA
Chapter 4

Prof. Dr. Renate Renkawitz-Pohl
Zell- und Entwicklungsbiologie
FB Biologie
Universität Marburg
Karl-von-Frisch-Straße
D-35032 Marburg
Germany
Chapter 1

Peter Rice
Lion Bioscience
Compass House
80–82 Newmarket Road
Cambridge CB5 8DZ
United Kingdom
Chapter 12

Dr. Stephen J. Robinson
AAFC, Saskatoon Research Center
Molecular Genetics Section
107 Science Place
Saskatoon, Saskatchewan, S7N 0X2
Canada
Chapter 5

Dr. Stephen C. Roemer
Advanced Research and Development
Li-Cor, Inc.
4308 Progressive Avenue
P.O. Box 4000
Lincoln, NE 68504
USA
Chapter 7

Dr. Kevin Rozwadowski
AAFC, Saskatoon Research Center
Molecular Genetics Section
107 Science Place
Saskatoon, Saskatchewan, S7N 0X2
Canada
Chapter 5

Dr. Stephen W. Scherer
The Hospital for Sick Children
HSC Research Institute
555 University Avenue
Toronto, Ontario, M5G 1X8
Canada
Chapter 2

Dr. Joseph D. Schrag
Biotechnology Research Institute
6100 Royalmont Avenue
Montreal, Quebec, H4P 2R2
Canada
Chapter 14

Dr. Christoph W. Sensen
University of Calgary
Faculty of Medicine
Department of Biochemistry and Molecular
Biology
Calgary, Alberta, T2N 4N1
Canada
Chapters 15 and 18

Dr. Andrew G. Sharpe
AAFC, Saskatoon Research Center
Molecular Genetics Section
107 Science Place
Saskatoon, Saskatchewan, S7N 0X2
Canada
Chapter 5

Dr. Daniel C. Tessier
National Research Council of Canada
6100 Royalmount Avenue
Montreal, Quebec, H4P 2R2
Canada
Chapter 8

Dr. David Y. Thomas
McGill University
Faculty of Medicine
Department of Biochemistry
McIntyre Medical Sciences Building
Room #801
3655 Promenade Sir William Osler
Montreal, Quebec, H3G 1Y6
Canada
Chapter 8

Dr. Theerayut Toojinda
DNA Fingerprinting Unit
Biotechnology Center
Kasetsart University
Kampangsaeen Campus
Nakorn Pathom 73140
Thailand
Chapter 6

Dr. Somvong Tragoonrung
DNA Fingerprinting Unit
Biotechnology Center
Kasetsart University
Kampangsaeen Campus
Nakorn Pathom 73140
Thailand
Chapter 6

Prof. Lap Chee Tsui
The Hospital for Sick Children
HSC Research Institute
555 University Avenue
Toronto, Ontario, M5G 1X8
Canada
Chapter 2

Dr. Apichart Vanavichit
DNA Fingerprinting Unit
Biotechnology Center
Kasetsart University
Kampangsaeen Campus
Nakorn Pathom 73140
Thailand
Chapter 6

XIV *Contributors*

Dr. David Wishart
University of Alberta
Faculty of Pharmacy and Pharmaceutical
Sciences
Dentistry/Pharmacy Centre 2123
Edmonton, Alberta, T6G 2N8
Canada
Chapter 13

Dr. Evgeni M. Zdobnov
EMBL Outstation
The European Bioinformatics Institute
Hinxton Hall, Hinxton
Cambridge, CB10 1SD
United Kingdom
Chapter 11

Dr. Zheru Zhang
Analytical R & D
Pharmaceutical Research Institute
Bristol – Myers Squibb Company
Syracuse, NY 13221-4755
USA
Chapter 10

Introduction

CHRISTOPH W. SENSEN

Calgary, Canada

Genomics has revolutionized biological and medical research and development over the last fifteen years. The speed and magnitude by which Genomics has outgrown the disciplines from which it originally developed have taken many by surprise. The rapid development of the field has left much of the early history of Genomics behind and many key events have not been recorded properly.

It may be even all but forgotten how the term “genomics” originated. According to the first editorial of *Genomics* (1987, **1**, 1–2), the term was coined by T. H. RODERICK from the Jackson Laboratories in Bar Harbor, MN, some time around 1987 in discussions with editors VICTOR A. MCKUSICK and FRANK H. RUDDLE, who were looking for suggestions to name their new journal.

There is no all-encompassing definition for genomics, the word is used with many meanings. At the time when MCKUSICK and RUDDLE wrote their editorial, they understood genomics to be mapping and sequencing to analyze the structure and organization of genomes. When the *Genomics* journal was founded, only three years had passed since the invention of automated DNA sequencers, which dominated the first phase of the development of genomics as a science. Thus a defi-

nition such as MCKUSICK and RUDDLE’s of the word genomics can be understood in the context of that time.

Today, genomics is very often subdivided into “structural genomics”, which deals with the determination of the complete sequence of genomes (DNA sequencing), or the complete set of proteins (proteome) in an organism (proteomics), and “functional genomics”, which studies the functioning of genes and metabolic pathways (metabolomics) or the gene expression patterns in an organism (chip technologies). To complicate matters, X-ray crystallographers have adopted the term structural genomics to refer to protein 3-D structure determination.

For the purpose of this book, genomics has the broader meaning of “genome research”, including bioinformatics and other studies of the genome and proteome to understand the blueprint and function of organisms. Many of the technologies that are part of today’s genomics toolkit were developed previously and then automated in an attempt to apply them in large-scale, high-throughput environments. Some people, including the late Canadian Nobel laureate MICHAEL SMITH (UBC), have claimed that they were doing genomics all along, which is true to a certain degree when

using a broad definition of genomics, considering its strong roots in molecular biology, biophysics, and biochemistry. With such a definition, we may say that genomics really started when WATSON and CRICK discovered the structure of DNA.

Without a doubt, the introduction of computers into molecular biology laboratories was one of the key factors in the development of genomics. Laboratory automation led to the production of large amounts of data, and the need to analyze, combine, and understand these resulted in the development of “bioinformatics”, a new discipline at the interface of several traditional disciplines. Bioinformatics is the glue that integrates all the diverse aspects of genomics. Of similar importance to the development of the field is the development of laser-based technologies. The use of laser-based systems, which can be coupled to computerized detection systems, has replaced most of the radioactive techniques in genomics laboratories, allowing the complete automation of many types of experiments.

Considering the rapid pace of development, it is quite difficult to organize a book that reflects all aspects of genomics. Chapters about model organisms are followed by overviews of the key technologies. Because of the importance to the field, several chapters are dedicat-

ed to bioinformatics. Genomics is a science with huge impact on society, thus ethical and legal issues that need to be dealt with arise daily. One of the book chapters is devoted to ethical and legal aspects of genome research. The book closes with an outlook to future developments in genomics.

With the completion of the human genome, the true tasks for genomics are only starting to emerge. We are far from understanding how organisms with small genomes function, let alone how the human genome is organized. As more and more scientific disciplines get “genomicized”, the field will undergo continual transformation, thus a book like this one can only capture a flavor and a moment in time. This may be frustrating to some, but this book is intended to summarize the essence of the first fifteen years of research and development in genomics, during which the cornerstone for a very exciting future was laid.

Calgary, January 2002 Christoph W. Sensen