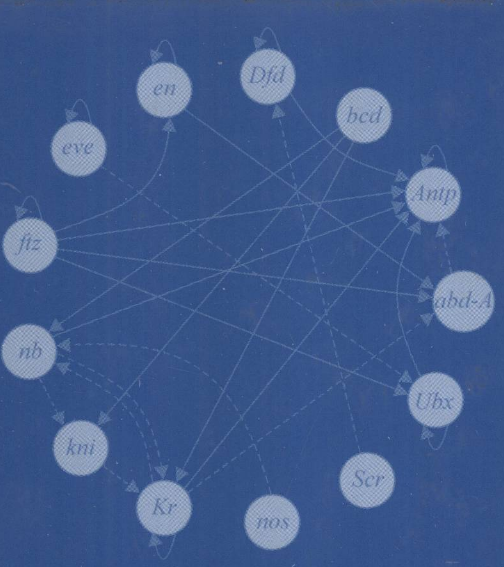

Bioinformatics of Genome Regulation and Structure II

Edited by
Nikolay Kolchanov
Ralf Hofestaedt
Luciano Milanesi

Springer

Q754-53
B615
2004

BIOINFORMATICS OF GENOME REGULATION AND STRUCTURE II

Edited by

NIKOLAY KOLCHANOV

Institute of Cytology & Genetics, Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia

RALF HOFESTAEDT

Bielefeld University, Bielefeld, Germany

LUCIANO MILANESI

CNR-ITB Institute of Biomedical Technologies, Segrate (Milano), Italy

Sponsored by the Institute of Cytology and Genetics, Siberian Branch of the Russian Academy of Sciences (Russia) and the Institute for Biomedical Technologies ITB-CNR (Italy)

华南理工大学
图书馆藏书

Springer

E200603334

Library of Congress Control Number: 2005933710

ISBN-10: 0-387-29450-3 e-ISBN-10: 0-387-29455-4
ISBN-13: 978-0-387-29540-6

Printed on acid-free paper.

© 2006 Springer Science+Business Media, Inc.
All rights reserved. This work may not be translated or copied in whole or in part without the written permission of the publisher (Springer Science+Business Media, Inc., 233 Spring Street, New York, NY 10013, USA), except for brief excerpts in connection with reviews or scholarly analysis. Use in connection with any form of information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed is forbidden.
The use in this publication of trade names, trademarks, service marks and similar terms, even if they are not identified as such, is not to be taken as an expression of opinion as to whether or not they are subject to proprietary rights.

Printed in the United States of America.

9 8 7 6 5 4 3 2 1 SPIN 11411925

springeronline.com

BIOINFORMATICS OF GENOME REGULATION AND STRUCTURE II

Contributing Authors

I. Abnizova
D.A. Afonnikov
A. Alexeevski
E. Ananko
P. Arrigo
A.S. Arseniev
T.V. Astakhova
G. Bazykin
S.M. Bendre
E. Beresikov
P.M. Beskaravainy
H. Binder
A. Blinov
R. te Boekhorst
A.S. Brok-Volchanski
T. Busygina
S. Cerutti
M.B. Chaley
C. Chaouiya
P. Chavatte
M. Chen
A. Chugunov
V.A. Debelov
A.A. Deev
P.S. Demenkov
G.V. Demidenko
R.G. Efremov
A. Ershova
S.I. Fadeev
M. Fattore
E.Ya. Frisman
D. Furman
M. Fursov
Yu.A. Gaidov
M.S. Gelfand
A.V. Gerasimova
W.R. Gilks
T. Gojobori
V.P. Golubyatnikov
V. Gor
A. Gorban
K. Gorbunov
D.A. Grigorovich
M. Heisler
R. Hofestaedt
E. Ignatieva
K. Ikeo
V.A. Ivanisenko
P. Jain
H. Jönsson
S.G. Kamzolova
A.A. Kanapin

此为试读, 需要完整PDF请访问: www.ertongbook.com

A. Karyagina
A. Katokhin
A. Katyshev
T.M. Khlebodarova
J. Ko
A.V. Kochetov
V.V. Kogai
N.A. Kolchanov
A. Kondrashov
Yu. Konstantinov
E.V. Korotkov
E.G. Kostyanicina
E.A. Kotelnikova
M.A. Krestyanova
N.A. Kudryashov
K. Kumar
V.A. Kuznetsov
O.N. Laikova
A.A. Laskin
V.G. Levitsky
V.A. Likhoshvai
N. Limova
E.K. Litvenko
I. Lokhova
V.I. Lukyanov
V. Lyubetsky
V.Y. Makeev
I.S. Masulis
I. Merelli
V. Merkulov
T. Merkulova
E.M. Meyerowitz
D. Miginsky
L. Milanesi
A.A. Mironov
V. Mironova
C.K. Mitra
E. Mjolsness
L.F. Murga
E. Myasnikova
E. Nedosekina
Y. Nishio
O. Novikova
A. Ogurtsov
N. Omelyanchuk
M.J. Ondrechen
Yu.L. Orlov
D. Oshchepkov
A.F. Osipov
A.A. Osypov
O.N. Ozoline
A. Palyanov
E.M. Panina
L. Pattini
S.V. Petrova
A.G. Pichueva
V.V. Pickalov
S.S. Pintus
O.A. Podkolodnaya
N. Podkolodny
A.A. Polyansky
A. Poplavsky
T. Popova
M.A. Pozdnyakov
A.L. Proscura
Yu.A. Purtov
S. Ramachandran
D.A. Ravcheev
G.V. Reddy
J. Reinitz
D.A. Rodionov
I.B. Rogozin
A. Romashchenko
M.A. Roytberg
L. Rusin
G. Sachdeva
A. Samsonova
M. Samsonova
L. Sánchez
K.V. Shaitan
B.E. Shapiro
K.S. Shavkunov
A.A. Shelenkov
T.I. Shipilov

K.G. Skryabin
O.G. Smirnova
A.A. Sorokin
S. Spirin
I. Stepanenko
S. Surkova
Y. Surya pavan
R.N. Tchuraev
K.B. Tereshkina
D. Thieffry
I. Titov
V. Trifonov
I.I. Tsitovich
V.P. Turutina
Y. Usuda
V. V'yugin
G. Vasiliev
O.V. Vishnevsky
A.G. Vitreschak
E.E. Vityaev
E.P. Volokitin
P.E. Volynsky
D. Vorobiev
K. Walter
Y. Wei
X. Xia
O.L. Zhdanova
A. Zinovyev

Preface

The last 15 years in development of biology were marked with accumulation of unprecedentedly huge arrays of experimental data. The information was amassed with exclusively high rates due to the advent of highly efficient experimental technologies that provided for high throughput genomic sequencing; of functional genomics technologies allowing investigation of expression dynamics of large groups of genes using expression DNA chips; of proteomics methods giving the possibility to analyze protein compositions of cells, tissues, and organs, assess the dynamics of the cell proteome, and reconstruct the networks of protein–protein interactions; and of metabolomics, in particular, high resolution mass spectrometry study of cell metabolites, and distribution of metabolic fluxes in the cells with a concurrent investigation of the dynamics of thousands metabolites in an individual cell.

Analysis, comprehension, and use of the tremendous volumes of experimental data reflecting the intricate processes underlying the functioning of molecular genetic systems are unfeasible in principle without the systems approach and involvement of the state-of-the-art information and computer technologies and efficient mathematical methods for data analysis and simulation of biological systems and processes.

The need in solving these problems initiated the birth of a new science—postgenomic bioinformatics or systems biology *in silico*. The following problems embody the key directions of the systems computer biology: (1) Integration of the heterogeneous experimental data that are obtained by various methods of structural and functional genomics, transcriptomics, proteomics, metabolomics, and other approaches of modern biology in databases; (2) Integrated computer analysis aimed at detection of the patterns in functioning of molecular genetic systems of the living organisms from microorganism to humans; (3) Construction of mathematical models of molecular biological and molecular genetic systems and processes that would form the background for analyzing the mechanisms involved in realization of genetic information; and (4) Investigation of the principles underlying organization and function of gene networks, which provide for formation of phenotypic characteristics of the living organisms basing on the information encoded in their genomes.

These issues of the utmost importance were discussed at the International Conference on Bioinformatics of Genome Regulation and Structure (BGRS), organized biennially by the Laboratory of Theoretical Genetics with the Institute of Cytology and Genetics, Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia. BGRS'2004, hold in July 2004, was the fourth event in the series (http://www.bionet.nsc.ru/meeting/bgrs2004/).

The program of the Fourth International Conference comprised five main sections, covering most topical aspects of theoretical, experimental, and applied directions of bioinformatics research, namely: (i) computer structural and functional genomics; (ii) computer structural and functional proteomics; (iii) computer evolutionary biology; (iv) computer systems biology, and (v) new approaches to analysis of data and processes in molecular biology.

We are glad to offer to the readers the second book *Bioinformatics of Genome Regulation and Structure,* which summarizes results of the works presented at BGRS'2004. The book contains selected reviewed papers of the Conference participants from many countries all over the world including Canada, China, France, Germany, India, Italy, Japan, The Netherlands, Russia, Singapore, Sweden, United Kingdom, USA.

We hope that this book will be useful for the scientists involved in basic and applied research in the field of experimental and theoretical studies of structure–function organization of genomes, on the one hand, and teachers and students of universities and colleges, mathematicians and biologists, on the other, that is, to a wide range of readers who are interested in the modern state, problems, possibilities, and prospects of bioinformatics.

Professor Nikolay Kolchanov
Head of the Laboratory of Theoretical Genetics
Institute of Cytology and Genetics, Siberian Branch
of the Russian Academy of Sciences, Novosibirsk, Russia
Chairman of the Conference

Professor Ralf Hofestaedt
Faculty of Technology, Bioinformatics Department
University of Bielefeld, Germany
Co-Chairman of the Conference

Contents

Contributing Authors xi

Preface xv

Part 1. COMPUTATIONAL STRUCTURAL, FUNCTIONAL AND EVOLUTIONARY GENOMICS 1

RECOGNITION OF CODING REGIONS IN GENOME ALIGNMENT
T.V. Astakhova, S.V. Petrova, I.I. Tsitovich, M.A. Roytberg 3

PREDICTING sRNA GENES IN THE GENOME OF *E. COLI*
BY THE PROMOTER-SEARCH ALGORITHM PlatProm
A.S. Brok-Volchanski, I.S. Masulis, K.S. Shavkunov, V.I. Lukyanov,
Yu.A. Purtov, E.G. Kostyanicina, A.A. Deev, O.N. Ozoline 11

CONTENT SENSORS BASED ON CODON STRUCTURE AND DNA
METHYLATION FOR GENE FINDING IN VERTEBRATE GENOMES
X. Xia 21

THE SITEGA TOOL FOR RECOGNITION AND CONTEXT ANALYSIS
OF TRANSCRIPTION FACTOR BINDING SITES: SIGNIFICANT
DINUCLEOTIDE FEATURES BESIDES THE CANONICAL
CONSENSUS EXEMPLIFIED BY SF-1 BINDING SITE
V. Levitsky, E. Ignatieva, G. Vasiliev, N. Limova, T. Busygina,
T. Merkulova, N. Kolchanov 31

TRANSCRIPTION REGULATORY REGIONS DATABASE (TRRD): A SOURCE OF EXPERIMENTALLY CONFIRMED DATA ON TRANSCRIPTION REGULATORY REGIONS OF EUKARYOTIC GENES
N. Kolchanov, E. Ignatieva, O. Podkolodnaya, E. Ananko, I. Stepanenko, T. Merkulova, T. Khlebodarova, V. Merkulov, N. Podkolodny, D. Grigorovich, A. Poplavsky, A. Romashchenko 43

ARTSITE DATABASE: COMPARISON OF *IN VITRO* SELECTED AND NATURAL BINDING SITES OF EUKARYOTIC TRANSCRIPTION FACTORS
T. Khlebodarova, O. Podkolodnaya, D. Oshchepkov, D. Miginsky, E. Ananko, E. Ignatieva 55

COMPARATIVE ANALYSIS OF ELECTROSTATIC PATTERNS FOR PROMOTER AND NONPROMOTER DNA IN *E. COLI* GENOME
S.G. Kamzolova, A.A. Sorokin, P.M. Beskaravainy, A.A. Osypov 67

ANALYSIS OF NUCLEOSOME FORMATION POTENTIAL AND CONFORMATIONAL PROPERTIES OF HUMAN J1-J2 AND D2-D1 TYPE ALPHA SATELLITE DNA
A. Katokhin, V. Levitsky, D. Oshchepkov, A. Poplavsky, V. Trifonov, D. Furman 75

VMM: A VARIABLE MEMORY MARKOV MODEL PREDICTION OF NUCLEOSOME FORMATION SITES
Yu.L. Orlov, V.G. Levitsky, O.G. Smirnova, O.A. Podkolodnaya, T.M. Khlebodarova, N.A. Kolchanov 85

DNA NUCLEOSOME ORGANIZATION OF THE PROMOTER GENE REGIONS
V.G. Levitsky, A.G. Pichueva, A.V. Kochetov, L. Milanesi 97

A TOOLBOX FOR ANALYSIS OF RNA SECONDARY STRUCTURE BASED ON GENETIC ALGORITHM
I. Titov, D. Vorobiev, A. Palyanov 105

COMPARATIVE GENOMICS AND EVOLUTION OF BACTERIAL REGULATORY SYSTEMS
M.S. Gelfand, A.V. Gerasimova, E.A. Kotelnikova, O.N. Laikova, V.Y. Makeev, A.A. Mironov, E.M. Panina, D.A. Ravcheev, D.A. Rodionov, A.G. Vitreschak 111

COMPARATIVE WHOLE GENOME SEQUENCE ANALYSIS OF CORYNEBACTERIA
Y. Nishio, Y. Usuda, T. Gojobori, K. Ikeo 121

此为试读,需要完整PDF请访问: www.ertongbook.com

NEW LTR RETROTRANSPOSABLE ELEMENTS FROM EUKARYOTIC GENOMES
O. Novikova, M. Fursov, E. Beresikov, A. Blinov 131

A GENOME-WIDE IDENTIFICATION OF MITOCHONDRIAL DNA TOPOISOMERASE I IN *ARABIDOPSIS*
A. Katyshev, I.B. Rogozin, Yu. Konstantinov 141

CHANGE IN CpG CONTEXT IS A LEADING CAUSE OF CORRELATION BETWEEN THE RATES OF NON-SYNONYMOUS AND SYNONYMOUS SUBSTITUTIONS IN RODENTS
G. Bazykin, A. Ogurtsov, A. Kondrashov 147

UNIVERSAL SEVEN–CLUSTER STRUCTURE OF GENOME FRAGMENT DISTRIBUTION: BASIC SYMMETRY IN TRIPLET FREQUENCIES
A. Gorban, A. Zinovyev, T. Popova 153

NEW METHODS TO INFER DNA FUNCTION FROM SEQUENCE INFORMATION
I. Abnizova, R. te Boekhorst, K. Walter, W.R. Gilks 165

REVELATION AND CLASSIFICATION OF DINUCLEOTIDE PERIODICITY OF BACTERIAL GENOMES USING THE METHOD OF INFORMATION DECOMPOSITION
A.A. Shelenkov, M.B. Chaley, K.G. Skryabin, E.V. Korotkov 179

ALGORITHMS TO RECONSTRUCT EVOLUTIONARY EVENTS AT MOLECULAR LEVEL AND INFER SPECIES PHYLOGENY
V. Lyubetsky, K. Gorbunov, L. Rusin, V. V'yugin 189

Part 2. COMPUTATIONAL STRUCTURAL AND FUNCTIONAL PROTEOMICS 205

MINING FROM COMPLETE PROTEOMES TO IDENTIFY ADHESINS AND ADHESIN-LIKE PROTEINS: A RAPID AID TO EXPERIMENTAL RESEARCHERS
S. Ramachandran, P. Jain, K. Kumar, G. Sachdeva 207

CENTRAL MOMENTS BASED STATISTICAL ANALYSIS FOR THE DETERMINATION OF FUNCTIONAL SITES IN PROTEINS WITH THEMATICS
L.F. Murga, J. Ko, Y. Wei, M.J. Ondrechen 215

AMINO ACIDS SURFACE PATTERNS IN PROTEIN DOMAIN FUNCTIONALITY ANALYSIS
I. Merelli, L. Pattini, S. Cerutti, L. Milanesi 225

COMPUTER SIMULATIONS OF ANIONIC UNSATURATED LIPID BILAYER—A SUITABLE MODEL TO STUDY MEMBRANE INTERACTIONS WITH A CELL-PENETRATING PEPTIDE
A.A. Polyansky, P.E. Volynsky, A.S. Arseniev, R.G. Efremov 235

THE ROLE OF WATER IN HOMEODOMAIN–DNA INTERACTION
A. Karyagina, A. Ershova, S. Spirin, A. Alexeevski 247

MOLECULAR MODELING OF HUMAN MT_1 AND MT_2 MELATONIN RECEPTORS
A. Chugunov, P. Chavatte, R. Efremov 259

MOLECULAR DYNAMICS OF SMALL PEPTIDES USING ERGODIC TRAJECTORIES
K.V. Shaitan, K.B. Tereshkina 271

A PERIODICAL NATURE OF 94 PROTEIN FAMILIES
V.P. Turutina, A.A. Laskin, N.A. Kudryashov, K.G. Skryabin, E.V. Korotkov 285

PREDICTION OF CONTACT NUMBERS OF AMINO ACID RESIDUES USING A NEURAL NETWORK MODEL
D.A. Afonnikov 297

FASTPROT: A COMPUTATIONAL WORKBENCH FOR RECOGNITION OF THE STRUCTURAL AND FUNCTIONAL DETERMINANTS IN PROTEIN TERTIARY STRUCTURES
V.A. Ivanisenko, S.S. Pintus, P.S. Demenkov, M.A. Krestyanova, E.K. Litvenko, D.A. Grigorovich, V.A. Debelov 305

A MARKOV MODEL FOR PROTEIN SEQUENCES
Y. Surya pavan, C.K. Mitra, S.M. Bendre 317

PROTEOME COMPLEXITY MEASURES BASED ON COUNTING OF DOMAIN-TO-PROTEIN LINKS FOR REPLICATIVE AND NON-REPLICATIVE DOMAINS
V.A. Kuznetsov, V.V. Pickalov, A.A. Kanapin 329

Part 3. COMPUTATIONAL SYSTEM BIOLOGY 343

A SOFTWARE ARCHITECTURE FOR DEVELOPMENTAL MODELING IN PLANTS: THE COMPUTABLE PLANT PROJECT
V. Gor, B.E. Shapiro, H. Jönsson, M. Heisler, G.V. Reddy, E.M. Meyerowitz, E. Mjolsness 345

PREDICTION AND ALIGNMENT OF METABOLIC PATHWAYS
M. Chen, R. Hofestaedt 355

GENERAL PRINCIPLES OF ORGANIZATION AND LAWS OF FUNCTIONING IN GOVERNING GENE NETWORKS
R.N. Tchuraev 367

FROM GRADIENTS TO STRIPES: A LOGICAL ANALYSIS OF *DROSOPHILA* SEGMENTATION GENETIC NETWORK
D. Thieffry, C. Chaouiya, L. Sánchez 379

SELF-OSCILLATIONS IN HYPOTHETICAL GENE NETWORKS
V.A. Likhoshvai, V.V. Kogai, S.I. Fadeev 391

PERIODIC TRAJECTORIES AND ANDRONOV–HOPF BIFURCATIONS IN MODELS OF GENE NETWORKS
V.P. Golubyatnikov, V.A. Likhoshvai, E.P. Volokitin, Yu.A. Gaidov, A.F. Osipov 405

ON THE PROBLEM OF SEARCH FOR STATIONARY POINTS IN REGULATORY CIRCUITS OF GENE NETWORKS
V.A. Likhoshvai 415

MODELING OF GENE EXPRESSION BY THE DELAY EQUATION
V.A. Likhoshvai, G.V. Demidenko, S.I. Fadeev 421

AGNS—A DATABASE ON EXPRESSION OF ARABIDOPSIS GENES
N. Omelyanchuk, V. Mironova, A. Poplavsky, N. Podkoldny, N. Kolchanov, E. Mjolsness, E. Meyerowitz 433

STUDY OF THE INTERACTIONS BETWEEN VIRAL AND HUMAN GENOMES DURING TRANSFORMATION OF B CELLS WITH EPSTEIN–BARR VIRUS
E. Ananko, D. Oshchepkov, E. Nedosekina, V. Levitsky, I. Lokhova, O. Smirnova, V. Likhoshvai, N. Kolchanov 443

PROBING GENE EXPRESSION: SEQUENCE-SPECIFIC HYBRIDIZATION ON MICROARRAYS
H. Binder 451

DETERMINATION OF THE DEVELOPMENTAL AGE OF A *DROSOPHILA* EMBRYO FROM CONFOCAL IMAGES OF ITS SEGMENTATION GENE EXPRESSION PATTERNS
E. Myasnikova, A. Samsonova, S. Surkova, M. Samsonova, J. Reinitz 467

Part 4. BIOMOLECULAR DATA AND PROCESSES ANALYSIS 479

TOPICAL CLUSTERING OF BIOMEDICAL ABSTRACTS BY SELF-ORGANIZING MAPS
M. Fattore, P. Arrigo 481

SOFTWARE FOR ANALYSIS OF GENE REGULATORY SEQUENCES BY KNOWLEDGE DISCOVERY METHODS
E.E. Vityaev, T.I. Shipilov, M.A. Pozdnyakov, O.V. Vishnevsky, A.L. Proscura, Yu.L. Orlov, P. Arrigo 491

A MATHEMATICAL MODEL OF THE DISCONTINUOUS GENETIC STRUCTURES FIXATION PROCESS
E.Ya. Frisman, O.L. Zhdanova 499

References 511

Index 553

PART 1

COMPUTATIONAL STRUCTURAL, FUNCTIONAL AND EVOLUTIONARY GENOMICS

此为试读,需要完整PDF请访问: www.ertongbook.com