

CONTENTS

Introduction

Bioinformatics: Mystery, Astrology or Service Technology?	1
<i>Frank Eisenhaber</i>	
Mystery	2
Astrology	6
Service Technology	7

Section I: Deriving Biological Function of Genome Information with Biomolecular Sequence and Structure Analysis

1. Reliable and Specific Protein Function Prediction by Combining Homology with Genomic(s) Context	13
<i>Martijn A. Huynen, Berend Snel and Toni Gabaldón</i>	
Types of Genomic Context	14
Accuracy and Genomic Coverage of Context	
Based Predictions	16
Experimentally Verified Context Predictions	17
Practical Examples of New Protein Function Predictions	
Based on Genomics Data	18
Discussion	25
2. Clues from Three-Dimensional Structure Analysis and Molecular Modelling: New Insights into Cytochrome P450 Mechanisms and Functions	30
<i>Karin Schleinkofer and Thomas Dandekar</i>	
Modelling	31
Molecular Dynamic Simulation	35
3. Prediction of Protein Function: Two Basic Concepts and One Practical Recipe	39
<i>Frank Eisenhaber</i>	
The Beginning: Deriving the Protein Sequence and the Definition of Protein Function	40
Concept No. 1: Function Inheritance from a Common Ancestor Gene	40
Limitations of the Homology Search Concept	41
Concept No. 2: Lexical Analysis, Physical Interpretation and Sequence Motif-Function Correlations	42
A Recipe for Analyzing Protein Sequences	43

Section II: Complementing Biomolecular Sequence Analysis with Text Mining in Scientific Articles

4. Extracting Information for Meaningful Function Inference through Text-Mining	57
<i>Hong Pan, Li Zuo, Rajaraman Kanagasabai, Zhuo Zhang, Vidhu Choudhary, Bijayalaxmi Mohanty, Sin Lam Tan, S.P.T. Krishnan, Pardha Sarathi Veladandi, Archana Meka, Weng Keong Choy, Sanjay Swarup and Vladimir B. Bajic</i>	
Scope and Nature of Text-Mining in Biomedical Domain	58
Focus of Our Text-Mining Systems	59
Supporting Text-Mining Systems for Gene Regulatory Networks Reconstruction	59
Dragon TF Relation Extractor (DTFRE)	60
Rulebase Construction	61
Rule Representation	62
Learning Algorithm	62
Mining Associations of Transcription Factors by Dragon TF Association Miner	63
Exploring Metabolome of <i>Arabidopsis thaliana</i> and Other Plant Species by Dragon Metabolome Explorer	65
Comparative Analysis of Bacterial Species	69
5. Literature and Genome Data Mining for Prioritizing Disease-Associated Genes	74
<i>Carolina Perez-Iratxeta, Peer Bork and Miguel A. Andrade</i>	
Mapping Symptoms to Gene Functions	75
Sequence Homology Based Searches	76
Performance of the System	77
Examples on How the System Works	78
Limitations, Scope, and Further Directions	79
The G2D Web Server	81

Section III: Mechanistic Predictions from the Analysis of Biomolecular Networks

6. Model-Based Inference of Transcriptional Regulatory Mechanisms from DNA Microarray Data	85
<i>Harman J. Bussemaker</i>	
Two Classes of Tools for Finding Motifs from Expression Data	86
REDUCE: Motif-Based Regression Analysis of the Transcriptome	86
From Central Dogma to "Omics Law"	87
Three Different Ways of Using Regression Analysis	88
MA-Networker: Integrating Occupancy and Transcriptome Data	89

7. The Predictive Power of Molecular Network Modelling: Case Studies of Predictions with Subsequent Experimental Verification	95
<i>Stefan Schuster, Edda Klipp and Marko Marhl</i>	
Optimal Time Course of Gene Expression	97
A Previously Unknown Metabolic Pathway	99
Decoding of Calcium Oscillations	100
Discussion	101

Section IV: Mechanistic Predictions from the Analysis of Biomolecular Sequence Populations: Considering Evolution for Function Prediction

8. Theory of Early Molecular Evolution: Predictions and Confirmations	107
<i>Edward N. Trifonov</i>	
Consensus Temporal Order of Amino Acids	109
Reconstruction of the Origin and Evolution of the Triplet Code	109
Prediction I. Early Proteins Were Glycine-Rich	110
Prediction II. The Earliest Protein Sequences Were a Mosaic of Two Independent Alphabets	112
Prediction III. Fundamental Binary Code of Protein Sequences	112
Prediction IV. Domestication of Life?	114
Linking the Codon Chronology with Other Events of Early Evolution	115
9. Hitchhiking Mapping: Limitations and Potential for the Identification of Ecologically Important Genes	117
<i>Christian Schlötterer</i>	
Microsatellites—A Widely Used Genetic Marker	118
The InRq Statistic	118
Mapping the Target of Selection	119
10. Understanding the Functional Importance of Human Single Nucleotide Polymorphisms	126
<i>Saurabh Asthana and Shamil Sunyaev</i>	
Comparative Sequence Analysis	127
Structure-Based Methods	128
Combined Methods	128
Methods of Validation	129
Detecting the Effects of Selection	129
Genome-Wide Analysis of Functional Polymorphic Variants	130
Significance for Medical Genetics	130



11. Correlations between Quantitative Measures of Genome Evolution, Expression and Function	133
<i>Yuri I. Wolf, Liran Carmel and Eugene V. Koonin</i>	
Evolution Rate, Expression Level and Expression Breadth	135
Evolution Rate, Gene Loss and Fitness Effect	135
Gene Duplications and Evolution Rate	137
Interactions between Three and More Parameters:	
More Than the Sum of the Parts?	138
The "Social Status" Model	138
Multi-Dimensional Structure of Expression, Evolution Rate, and Gene Loss Data	140
Appendix 1. Mutual Entropy of Gene Knockout Data and Phyletic Patterns	142
Appendix 2. Expression Level, Evolution Rate, and Gene Loss as Predictors of Viability of Gene Knockout Mutants	142
Index	145

