

Contents

CONTRIBUTORS.....	xiii
PREFACE	xv

CHAPTER 1	Human Genome Informatics: Coming of Age	1
	Christophe G. Lambert, Darrol J. Baker, George P. Patrinos	
1.1	Introduction	1
1.2	From Informatics to Bioinformatics and Genome Informatics	3
1.3	Informatics in Genomics Research and Clinical Applications	6
	1.3.1 Genome Informatics Analysis.....	6
	1.3.2 Genomics Data Sharing	7
	1.3.3 Genomic Variant Reporting and Annotation Tools.....	8
1.4	Pharmacogenomics and Genome Informatics.....	8
1.5	Databases, Artificial Intelligence, and Big-Data in Genomics.....	9
1.6	Conclusions.....	10
	Acknowledgments	10
	References.....	10

PART 1 Human Genome Informatics Applications

CHAPTER 2	Creating Transparent and Reproducible Pipelines: Best Practices for Tools, Data, and Workflow Management Systems.....	15
	Alexandros Kanterakis, George Potamias, Morris A. Swertz, George P. Patrinos	
2.1	Introduction.....	15
2.2	Existing Workflow Environment	16

2.3	What Software Should Be Part of a Scientific Workflow?	18
2.4	Preparing Data for Automatic Workflow Analysis	21
2.5	Quality Criteria for Modern Workflow Environments	22
2.5.1	Being Able to Embed and to Be Embedded	22
2.5.2	Support Ontologies	23
2.5.3	Support Virtualization	23
2.5.4	Offer Easy Access to Commonly Used Datasets	24
2.5.5	Support and Standardize Data Visualization	25
2.5.6	Enable "Batteries Included" Workflow Environments	26
2.5.7	Facilitate Data Integration, Both for Import and Export	27
2.5.8	Offer Gateways for High Performance Computing Environments	28
2.5.9	Engage Users in Collaborative Experimentation and Scientific Authoring	29
2.6	Benefits From Integrated Workflow Analysis in Bioinformatics	30
2.6.1	Enable Meta-Studies, Combine Datasets, and Increase Statistical Power	30
2.6.2	Include Methods and Data From Other Research Disciplines	30
2.6.3	Fight the Reproducibility Crisis	31
2.6.4	Spread of Open Source Policies in Genetics and Privacy Protection	31
2.6.5	Help Clinical Genetics Research	32
2.7	Discussion	33
	References	33
CHAPTER 3		
	How Cytogenetics Paradigms Shape Decision Making in Translational Genomics	45
	Christophe G. Lambert	
3.1	Introduction	45
3.2	Clinical Cytogenetic Testing	45
3.2.1	Karyotyping	45
3.2.2	Chromosomal Microarrays	47
3.3	From Cytogenetics to Cytogenomics in the Era of Next-Generation Sequencing	56
3.4	Conclusions	57
	References	58

CHAPTER 4	An Introduction to Tools, Databases, and Practical Guidelines for NGS Data Analysis 61
	Alexandros Kanterakis, George Potamias, George P. Patrinos
4.1	Introduction 61
4.2	Data Formats 62
4.3	Data Sources 64
4.4	Variant Data 66
4.5	NGS Pipelines 67
4.5.1	Read Alignment 67
4.5.2	Variant Calling 69
4.5.3	Downstream Analysis 71
4.5.4	RNA-seq 72
4.5.5	ChIP-Seq 77
4.6	Discussion 78
	References 80
CHAPTER 5	Proteomics and Metabolomics Data Analysis for Translational Medicine 91
	Theodora Katsila, George P. Patrinos
5.1	Introduction 91
5.2	The Need to Bridge the Gaps in the Era of Precision Medicine 92
5.3	Clinical Proteomics 93
5.3.1	The Power of the Proteome 93
5.3.2	Steps Prior to Routine Proteome Analyses 96
5.4	Clinical Metabolomics 97
5.4.1	The Advances and Promises of Metabolome 97
5.4.2	Needs Prior to Routine Metabolome Analyses 98
5.5	Computational and Chemoinformatic Tools 98
5.6	Strategies to Address Data Complexity 99
5.7	From Translational Medicine Data to Theranostics 101
5.8	Conclusions 103
	Acknowledgments 103
	References 103
CHAPTER 6	Incentives for Human Genome Variation Data Sharing 109
	George P. Patrinos
6.1	Introduction 109
6.2	Database Projects Linked to Scientific Journals 110
6.3	Microattribution and Nanopublication: An Innovative Publication Modality 112

6.3.1	The Concept of Microattribution.....	113
6.3.2	The Microattribution Process.....	116
6.3.3	Implementation of Microattribution.....	117
6.4	Microattribution: Hurdles From Concept to Implementation.....	119
6.5	Proposed Measures and Steps Forward.....	122
6.6	Conclusions.....	125
	References.....	127

PART 2 Human Genome Informatics Tools and Related Resources

CHAPTER 7	A Review of Tools to Automatically Infer Chromosomal Positions From dbSNP and HGVS Genetic Variants.....	133
	Alexandros Kanterakis, Theodora Katsila, George Potamias, George P. Patrinos, Morris A. Swertz	
7.1	Introduction.....	133
7.2	Existing Tools for HGVS Position Resolution.....	137
7.2.1	Mutalyzer.....	137
7.2.2	The HGVS Python Package.....	138
7.2.3	Variant Effect Predictor.....	139
7.2.4	Variation Reporter.....	140
7.2.5	Transvar.....	142
7.2.6	BioPython.....	143
7.3	The MutationInfo Pipeline.....	145
7.4	Results.....	148
7.4.1	Analysis of dbSNP Variants.....	149
7.4.2	Analysis of HGVS Variants.....	151
7.5	Discussion.....	152
7.6	Conclusions.....	152
	References.....	154

CHAPTER 8	Translating Genomic Information to Rationalize Drug Use.....	157
	Alexandros Kanterakis, Theodora Katsila, George P. Patrinos	
8.1	Introduction.....	157
8.2	Personalized PGx Profiling Using Whole Genome Sequencing.....	159
8.3	Towards Pharmacogenomic Data Integration.....	161
8.3.1	The Concept of Integrated PGx Assistant Services.....	162

8.3.2	Development of an Electronic PGx Assistant	163
8.4	Personalized PGx Translation Services	166
8.5	The Electronic PGx Assistant	168
8.5.1	Explore Service	169
8.5.2	Translation Service	171
8.5.3	Update Services	171
8.6	Translating PGx Knowledge Into Clinical Decision-Making	173
8.7	Conclusion and Future Perspectives	175
	Acknowledgments	176
	References	176

CHAPTER 9

	Minimum Information Required for Pharmacogenomics Experiments	179
	J. Kumuthini, L. Zass, Emilson Chimusa, Melek Chaouch, Collen Masimiremwa	
9.1	Pharmacogenomics	179
9.2	Data Standardization	180
9.3	Minimum Information Required for a DMET Experiment	182
9.3.1	Background	182
9.3.2	DMET Console Software Analysis	183
9.3.3	MIDE	184
9.3.4	Discussion	185
9.4	Pharmacogenomics Standardization: Challenges	189
9.5	Conclusions	190
	References	191
	Further Reading	193

CHAPTER 10

	Human Genomic Databases in Translational Medicine	195
	Theodora Katsila, Emmanouil Viennas, Marina Bartsakoulia, Aggeliki Komianou, Konstantinos Sarris, Giannis Tzimas, George P. Patrinos	
10.1	Introduction	195
10.2	Historical Overview of Genomic Databases	196
10.3	Genomic Database Types	196
10.4	Models for Database Management	199
10.5	General Variation Databases: Documentation of Variants of Clinical Significance	200
10.5.1	ClinVar Database	200
10.5.2	Data Sharing in ClinVar	203

10.6	Locus-Specific Databases in Translational Medicine.....	203
10.6.1	Comparison Among Various LSDBs.....	203
10.6.2	Identification of Causative Genomic Variants	204
10.6.3	Linking Genotype Information With Phenotypic Patterns.....	205
10.6.4	Selection of the Optimal Variant Allele Detection Strategy	207
10.7	National/Ethnic Genomic Databases: Archiving the Genomic Basis of Human Disorders on a Population-Specific Basis.....	208
10.8	NEGDBs in a Molecular Diagnostics Setting	212
10.8.1	NEGDBs and Society.....	213
10.9	Database Management Systems for LSDBs and NEGDBs.....	214
10.10	Future Challenges	216
10.11	Conclusions	218
	Acknowledgments	218
	References.....	218

CHAPTER 11 Artificial Intelligence: The Future Landscape of Genomic Medical Diagnosis: Dataset, In Silico Artificial Intelligent Clinical Information, and Machine Learning Systems.....

Darrol J. Baker

11.1	Introduction.....	223
11.2	What Is Artificial Intelligence and Machine Learning?...224	
11.2.1	Clinical Genomics Medical AI	224
11.2.2	The Emerging Role of AI	225
11.3	Artificial Intelligence: The AI, The Whole AI, and Nothing but the AI.....	225
11.3.1	Building an Artificial Intelligence Framework (AIF)	226
11.3.2	Security and Quantum Computing (QComp)	229
11.3.3	Complex Knowledge Management Systems I ...	230
11.3.4	Epochs Intelligence for Clinical Diagnostics Understanding.....	236
11.3.5	Clinical Patient Understanding	240
11.3.6	Epoch Deep Belief Neural Networks	241
11.3.7	Educational Aside	244
11.3.8	Extracting Information From the Medical Record	245
11.3.9	Eliciting Information From Experts.....	246

11.3.10 The Next Steps in Clinical AI.....	248
11.4 Connecting Artificial Intelligence and Machine Learning Metadata Analysis for Clinical Diagnostic Discovery	248
11.5 Conclusions.....	261
Acknowledgments	262
References.....	262
Further Reading.....	267
CHAPTER 12 Genomics England: The Future of Genomic Medical Diagnosis: Governmental Scale Clinical Sequencing and Potential Walled-Garden Impact on Global Data Sharing.....	269
Darrol J. Baker	
12.1 Introduction.....	269
12.2 Understanding the Genomics England Approach to Clinical Genomic Discovery.....	271
12.3 Conclusions.....	290
Acknowledgments	290
References.....	290
Further Reading.....	292
INDEX.....	293