

1. Record Nr.	UNINA9910830623303321
Titolo	Biological Data Integration : Computer and Statistical Approaches // coordinated by Christine Froidevaux, Marie-Laure Martin-Magniette, and Guillem Rigail
Pubbl/distr/stampa	London, England : , : ISTE Ltd, , [2023] ©2023
ISBN	1-394-25731-7 1-394-25729-5
Edizione	[First edition.]
Descrizione fisica	1 online resource (276 pages)
Disciplina	570.113
Soggetti	Systems biology
Lingua di pubblicazione	Inglese
Formato	Materiale a stampa
Livello bibliografico	Monografia
Nota di bibliografia	Includes bibliographical references and index.
Nota di contenuto	Cover -- Title Page -- Copyright Page -- Contents -- Preface -- Part 1. Knowledge Integration -- Chapter 1. Clinical Data Warehouses -- 1.1. Introduction to clinical information systems and biomedical warehousing: data warehouses for what purposes? -- 1.1.1. Warehouse history -- 1.1.2. Using data warehouses today -- 1.2. Challenge: widely scattered data -- 1.3. Data warehouses and clinical data -- 1.3.1. Warehouse structures -- 1.3.2. Warehouse construction and supply -- 1.3.3. Uses -- 1.4. Warehouses and omics data: challenges -- 1.4.1. Challenges of data volumetry and structuring omic data -- 1.4.2. Attempted solutions -- 1.5. Challenges and prospects -- 1.5.1. Toward general-purpose warehouses -- 1.5.2. Ethical dimension of the implementation and the use of warehouses -- 1.5.3. Origin and reproducibility -- 1.5.4. Data quality -- 1.5.5. Data warehousing federation and data sharing -- 1.6. References -- Chapter 2. Semantic Web Methods for Data Integration in Life Sciences -- 2.1. Data-related requirements in life sciences -- 2.1.1. Databases for the life sciences -- 2.1.2. Requirements -- 2.1.3. Common approaches: InterMine and BioMart -- 2.2. Semantic Web -- 2.2.1. Techniques -- 2.2.2. Implementation -- 2.3. Perspectives -- 2.3.1. Facilitating appropriation to users -- 2.3.2. Facilitating the appropriation by software programs: FAIR data -- 2.3.3. Federated queries -- 2.4. Conclusion -- 2.5.

References -- Chapter 3. Workflows for Bioinformatics Data Integration -- 3.1. Introduction -- 3.2. Bioinformatics data processing chains: difficulties -- 3.2.1. Designing a data processing chain -- 3.2.2. Analysis execution and reproducibility -- 3.2.3. Maintenance, sharing and reuse -- 3.3. Solutions provided by scientific workflow systems -- 3.3.1. Fundamentals of workflow systems -- 3.3.2. Workflow systems -- 3.4. Use case: RNA-seq data analysis. 3.4.1. Study description -- 3.4.2. From data processing chain to workflows -- 3.4.3. Data processing chains implemented as workflows: conclusion -- 3.5. Challenges, open problems and research opportunities -- 3.5.1. Formalizing workflow development -- 3.5.2. Workflow testing -- 3.5.3. Discovering and sharing workflows -- 3.6. Conclusion -- 3.7. References -- Part 2. Integration and Statistics -- Chapter 4. Variable Selection in the General Linear Model: Application to Multiomic Approaches for the Study of Seed Quality -- 4.1. Introduction -- 4.2. Methodology -- 4.2.1. Estimation of the covariance matrix $f^{\circ}q$ -- 4.2.2. Estimation of B -- 4.3. Numerical experiments -- 4.3.1. Statistical performance -- 4.3.2. Numerical performance -- 4.4. Application to the study of seed quality -- 4.4.1. Metabolomics data -- 4.4.2. Proteomics data -- 4.5. Conclusion -- 4.6. Appendices -- 4.6.1. Example of using the package MultiVarSel for metabolomic data analysis -- 4.6.2. Example of using the package MultiVarSel for proteomic data analysis -- 4.7. Acknowledgments -- 4.8. References -- Chapter 5. Structured Compression of Genetic Information and Genome-Wide Association Study by Additive Models -- 5.1. Genome-wide association studies -- 5.1.1. Introduction to genetic mapping and linkage analysis -- 5.1.2. Principles of genome-wide association studies -- 5.1.3. Single nucleotide polymorphism -- 5.1.4. Disease penetrance and odds ratio -- 5.1.5. Single marker analysis -- 5.1.6. Multi-marker analysis -- 5.2. Structured compression and association study -- 5.2.1. Context -- 5.2.2. New structured compression approach -- 5.3. Application to ankylosing spondylitis (AS) -- 5.3.1. Data -- 5.3.2. Predictive power evaluation -- 5.3.3. Manhattan diagram -- 5.3.4. Estimation for the most significant SNP aggregates -- 5.4. Conclusion -- 5.5. References -- Chapter 6. Kernels for Omics. 6.1. Introduction -- 6.2. Relational data -- 6.2.1. Data described by the kernel -- 6.2.2. Data described by a general (dis)similarity measure -- 6.3. Exploratory analysis for relational data -- 6.3.1. Kernel clustering -- 6.3.2. Kernel principal component analysis -- 6.3.3. Kernel self-organizing maps -- 6.3.4. Limitations of relational methods -- 6.4. Combining relational data -- 6.4.1. Data integration in systems biology -- 6.4.2. Kernel approaches in data integration -- 6.4.3. A consensual kernel -- 6.4.4. A parsimonious kernel that preserves the topology of the initial data -- 6.4.5. A complete kernel preserving the topology of the initial data -- 6.5. Application -- 6.5.1. Loading Tara Ocean data -- 6.5.2. Data integration by kernel approaches -- 6.5.3. Exploratory analysis: kernel PCA -- 6.6. Session information for the results of the example -- 6.7. References -- Chapter 7. Multivariate Models for Data Integration and Biomarker Selection in 'Omics Data -- 7.1. Introduction -- 7.2. Background -- 7.2.1. Mathematical notations -- 7.2.2. Terminology -- 7.2.3. Multivariate projection-based approaches -- 7.2.4. A criterion to maximize specific to each methodology -- 7.2.5. A linear combination of variables to reduce the dimension of the data -- 7.2.6. Identifying a subset of relevant molecular features -- 7.2.7. Summary -- 7.3. From the biological question to the statistical analysis -- 7.3.1. Exploration of one dataset: PCA -- 7.3.2. Classify samples: projection to latent structure discriminant analysis -- 7.3.3. Integration of two datasets: projection to latent structure and related methods --

7.3.4. Integration of several datasets: multi-block approaches -- 7.4. Graphical outputs -- 7.4.1. Individual plots -- 7.4.2. Variable plots -- 7.5. Overall summary -- 7.6. Liver toxicity study -- 7.6.1. The datasets. 7.6.2. Biological questions and statistical methods -- 7.6.3. Single dataset analysis -- 7.6.4. Integrative analysis -- 7.7. Conclusion -- 7.8. Acknowledgments -- 7.9. Appendix: reproducible R code -- 7.9.1. Toy examples -- 7.9.2. Liver toxicity -- 7.10. References -- List of Authors -- Index -- EULA.
