

1. Record Nr.	UNINA9910143748003321
Titolo	Database annotation in molecular biology // editor, Arthur M. Lesk
Pubbl/distr/stampa	Chichester, West Sussex ; ; Hoboken, NJ, : John Wiley, c2005
ISBN	1-280-23872-0 9786610238729 0-470-01242-0 0-470-85685-8
Descrizione fisica	1 online resource (267 p.)
Altri autori (Persone)	LeskArthur M
Disciplina	572.8
Soggetti	Bioinformatics Nucleotide sequence - Data processing Amino acid sequence - Data processing
Lingua di pubblicazione	Inglese
Formato	Materiale a stampa
Livello bibliografico	Monografia
Note generali	Description based upon print version of record.
Nota di bibliografia	Includes bibliographical references and index.
Nota di contenuto	Database Annotation in Molecular Biology; Contents; Preface; List of Contributors; 1 Annotation and Databases: Status and Prospects; 1.1 Introduction; 1.2 Annotation of Genomic Data; 1.3 Databases: Concepts and Definitions; 1.4 Access to Annotation Databases; Glossary; References; I THE DATABANKS; 2 Survey of Sequence Databases: Archival Projects; 2.1 Introduction; 2.2 Nucleotide Sequence Databases; 2.3 Swiss-Prot; 2.4 TrEMBL; 2.5 PIR; 2.6 UniProt; References; 3 Survey of Sequence Databases: Derived Databases; 3.1 Introduction; 3.2 Protein and Gene Family Databases; 3.3 Discussion; References 4 Databanks of Macromolecular Structure4.1 Introduction; 4.2 Background; 4.3 Archival Structural Databases Now; 4.4 Contextual Databases; 4.5 Derived Structural Data Databases; 4.6 Summary and View of the Future; References; 5 Gene Expression Databases; 5.1 Introduction; 5.2 What Do We Mean by Microarray Gene Expression Data?; 5.3 Data Complexity; 5.4 Minimum Information About a Microarray Experiment (MIAME); 5.5 Journals and MIAME; 5.6 Storage and Exchange Formats: MAGE-OM and MAGE-ML; 5.7 ArrayExpress; 5.8 Annotation Tools; 5.9 Curation; 5.10 Standardization and Semantics 5.11 Public Microarray Databases5.12 ArrayExpress, an Example of a

Public Repository; 5.13 Submissions to ArrayExpress; 5.14 MIAMExpress and Other MIAME Compliant Annotation Systems; 5.15 Databases of Protein Expression Patterns; 5.16 The Gene Expression Database (GXD); 5.17 Conclusion; References; II THE BASIS OF ANNOTATION; 6 Taxonomy: a Moving Target for Sequence Data; 6.1 Introduction; 6.2 Nomenclature; 6.3 Operational Definitions; 6.4 Searching for the Taxonomic Gold Standard; 6.5 Conclusions; References; 7 Genomics and Proteomics: Design and Sources of Annotation
7.1 Beyond the Sequence: the Challenge of Complete Genome Analysis
7.2 Extracting the Genes; 7.3 Organism Specific Peculiarities; 7.4 Topology of Genomes; 7.5 Gene Extraction Pipelines; 7.6 Added Value and Knowledge; 7.7 Beyond the Parts List; References; 8 Annotation of Protein Sequences; 8.1 Introduction; 8.2 What is Annotation?; 8.3 UniProt: Universal Protein Resource; 8.4 Protein Family Classification; 8.5 InterPro: Integrated Resource of Protein Families, Domains and Sites; 8.6 PIR Protein Families and Superfamilies; 8.7 Ontologies
8.8 Protein Names, Source Information and Unique Identifiers
8.9 Common Identification Errors; 8.10 Evidence Attribution; 8.11 Position Specific Annotations; 8.12 Rule-based Annotation; 8.13 Conclusions; Acknowledgements; References; 9 Issues in the Annotation of Protein Structures; 9.1 Data Harvesting; 9.2 Identification of the Biologically Relevant Assembly; 9.3 Taxonomy; 9.4 Sequence Recognition and Cross-reference; 9.5 Recognition of Secondary Structure Elements; 9.6 Validation of Structures; 9.7 Residue Identification; 9.8 Hetgroup Identification; 9.9 Solvent Handling
9.10 Miscellaneous Annotation Issues

Sommario/riassunto

Two factors dominate current molecular biology: the amount of raw data is increasing very rapidly and successful applications in biomedical research require carefully curated and annotated databases. The quality of the experimental data -- especially nucleic acid sequences -- is satisfactory; however, annotations depend on features inferred from the data rather than measured directly, for instance the identification of genes in genome sequences. It is essential that these inferences are as accurate as possible and this requires human intervention. With the recognition of the importance
