

Introduction to Advanced Bioinformatics

Introduction to Advanced Bioinformatics

Graduate-level overview – concise, organized, and ready for quick reference.

1. What Is Advanced Bioinformatics?

Aspect	Description
Definition	Application of sophisticated computational, statistical, and mathematical methods to analyze large-scale biological data (genomics, transcriptomics, proteomics, epigenomics, etc.).
Goal	Extract biologically meaningful patterns, build predictive models, and generate testable hypotheses that go beyond routine sequence analysis.
Key Distinction	Moves from “data handling” to integrative analytics , machine learning, network inference, and systems-level interpretation.

2. Core Areas & Typical Workflows

2.1. High-Throughput Sequencing (HTS) Data

- Raw data acquisition** – Illumina, PacBio, Oxford Nanopore reads.
- Quality control** – FastQC, MultiQC, adapter trimming (Trimmomatic, cutadapt).
- Alignment / assembly** – BWA-MEM, minimap2, STAR (RNA-seq), SPAdes, Flye.
- Variant calling & annotation** – GATK, FreeBayes, bcftools; annotation with VEP, ANNOVAR.
- Downstream analyses** – differential expression (DESeq2, edgeR), allele-specific expression, isoform reconstruction.

2.2. Multi-Omics Integration

Modality	Typical Data Type	Integration Strategies
Genomics	SNPs, CNVs, structural variants	eQTL mapping, GWAS meta-analysis
Transcriptomics	RNA-seq, single-cell RNA-seq	Co-expression networks, pseudo-time analysis

Introduction to Advanced Bioinformatics

Modality	Typical Data Type	Integration Strategies
Proteomics	Mass?spec peptide quantification	Correlation with transcript levels, PTM mapping
Epigenomics	ATAC?seq, ChIP?seq, DNA methylation	Chromatin state modeling, regulatory element annotation
Metabolomics	LC?MS, NMR spectra	Pathway enrichment, flux balance analysis

Integration frameworks: MOFA+, iCluster, mixOmics, Bayesian hierarchical models, deep learning (autoencoders, graph neural networks).

2.3. Systems Biology & Network Analysis

- **Gene?regulatory networks** – ARACNe, GENIE3, Inferelator.
- **Protein?protein interaction (PPI) networks** – STRING, BioGRID, network propagation.
- **Metabolic networks** – COBRA Toolbox, flux balance analysis (FBA).
- **Multi?layer (heterogeneous) networks** – multiplex graphs, network alignment.

2.4. Machine Learning & AI in Bioinformatics

Technique	Typical Application
Supervised learning (Random Forest, SVM, XGBoost)	Variant pathogenicity prediction, phenotype classification.
Deep learning (CNN, RNN, Transformer)	Motif discovery, splice site prediction, protein structure (AlphaFold).
Unsupervised learning (k?means, hierarchical clustering, t?SNE, UMAP)	Cell?type identification in scRNA?seq, dimensionality reduction.
Reinforcement learning	Adaptive experimental design, drug?target discovery.

Key libraries: scikit?learn, TensorFlow/PyTorch, Keras, Hugging Face Transformers, PyTorch Geometric.

3.