Bioinformatics Skill Development Workshop

From Programming to NGS Analysis

Section : I Data Analysis Using R and R-Studio

Modules at a Glance

- Module 0: Introduction to R and RStudio - Setup, data structures, R Markdown basics
- Module 1: Data Import, Cleaning and Transformation - CSV/Excel import, NA handling, dplyr, tidyr, joins
- Module 2: Descriptive Statistics - Mean, median, SD, IQR, skewness, kurtosis, frequency tables
- Module 3: Data Visualization - Foundations - ggplot2, histograms, box plots, bar charts, scatter plots, faceting
- Module 4: Data Visualization - Advanced - Violin plots, heatmaps, bubble plots, correlation plots, Plotly interactive charts
- Module 5: Probability and Distributions - Normal, binomial, Poisson; Shapiro-Wilk, Q-Q plots, Central Limit Theorem
- Module 6: Hypothesis Testing - t-tests, Wilcoxon, Mann-Whitney, Chi-square, confidence intervals, effect sizes
- Module 7: ANOVA and Post-Hoc Analysis - One-way and two-way ANOVA, Tukey HSD, DMRT, Bonferroni, Kruskal-Wallis
- Module 8: Correlation and Regression - Pearson/Spearman correlation, simple and multiple linear regression, diagnostics, logistic regression
- Module 9: Multivariate Analysis - PCA, scree plot, biplot, hierarchical and k-means clustering

What You Will Learn

- Import, clean, and reshape real-world datasets using the tidyverse
- Compute and interpret descriptive statistics for any dataset
- Assess normality and choose the right parametric or non-parametric test
- Perform t-tests, chi-square tests, one-way and two-way ANOVA
- Apply correct post-hoc tests: Tukey HSD, DMRT, and Bonferroni
- Build simple and multiple linear regression models and check assumptions
- Run PCA and cluster analysis for multivariate exploration
- Create publication-ready plots and interactive visualizations

Key R Packages Covered

- Data: dplyr, tidyr, readr, readxl, lubridate, janitor
- Visualization: ggplot2, plotly, pheatmap, GGally, corrplot
- Statistics: rstatix, car, ez, agricolae, MASS, e1071
- Multivariate: factoextra, FactoMineR, cluster

Section : II Python & Biopython

From Zero Programming to Structural Bioinformatics And Molecular Docking

What You Will Learn

Module- I : Fundamentals of Python

1. Python Foundations for Biology

Build a strong programming base using biological examples.
Learn how to store, manipulate, and process DNA, RNA, and protein data using Python.

2. Sequence Manipulation & String Processing

Work directly with biological sequences using Python string operations.
Perform GC content analysis, sequence transformations, and pattern detection.

3. Data Structures for Biological Data

Organize biological datasets using lists, tuples, and dictionaries.
Model real biological systems such as codon tables and amino acid properties.

4. Logical Thinking & Automation

Write efficient code using loops and conditional logic.
Automate repetitive biological analyses and filtering tasks.

5. Functions & Reusable Bioinformatics Code

Build reusable tools for sequence analysis such as GC calculation and translation.
Write modular and well-documented code.

6. File Handling & Scientific Computing

Work with FASTA files and biological datasets.
Use NumPy for numerical analysis and Matplotlib for visualization.

Module -II : Biopython & Core Bioinformatics Workflows

7. Biopython for Sequence Analysis

Use Biopython to manipulate DNA, RNA, and protein sequences.
Perform transcription, translation, and sequence transformations.

8. Working with Biological File Formats

Read and write FASTA and GenBank files.
Extract annotations and biological features programmatically.

9. NCBI Database Integration

Fetch sequences and biological records directly from NCBI databases.
Automate retrieval of genomic and literature data.

10. BLAST & Homology Analysis

Perform sequence similarity searches using BLAST.
Parse results and identify homologous sequences.

11. Multiple Sequence Alignment & Phylogenetics

Align sequences and construct phylogenetic trees.
Understand evolutionary relationships and sequence conservation.

12. Motif Analysis & Restriction Enzymes

Identify sequence motifs and restriction enzyme sites.
Analyze biological patterns within DNA sequences.

13. Structural Bioinformatics (PDB Analysis)

Parse protein structures and extract atomic and residue-level information.
Analyze structural properties such as B-factors and hydrophobicity.

Module -III : Advanced Applications in Bioinformatics

14. Structural Comparison & RMSD Analysis

Compare protein structures using superimposition techniques.
Understand structural similarity through RMSD metrics.

15. Network Biology & Sequence Similarity Networks

Convert BLAST results into network graphs.
Visualize relationships between proteins using interactive tools.

16. Molecular Docking & Drug Discovery

Understand the fundamentals of molecular docking.
Prepare proteins and ligands, run docking simulations, and interpret results.

17. Data Analysis with Pandas

Work with structured biological datasets.
Filter, group, and analyze protein and sequence data efficiently.

18. Dimensionality Reduction & Clustering

Apply PCA and clustering techniques to biological datasets.
Identify patterns and group similar biological entities.

19. Pipeline Automation & Workflow Design

Build end-to-end bioinformatics pipelines.
Automate data retrieval, analysis, and visualization steps.

Capstone Project

Bring everything together by performing a complete structural and sequence analysis of a target protein.
Integrate sequence retrieval, BLAST analysis, structural parsing, and visualization into a single workflow.

Section III : Next-Generation Sequencing (NGS) Analysis

This section provides a comprehensive introduction to NGS data analysis, covering essential computational skills, biological data handling, and practical workflows. Participants will gain hands-on experience using Linux-based environments and widely used bioinformatics tools for real-world genomic data analysis.

◇ Module 0 – Linux on Windows (WSL)

This module introduces the Linux environment, which is fundamental for bioinformatics workflows. Participants will understand why most bioinformatics tools are optimized for Linux systems and learn how to set up a Linux environment on Windows using Windows Subsystem for Linux (WSL).

- 0.1 Introduction to Linux and its importance in bioinformatics
- 0.2 Limitations of running bioinformatics tools on native Windows
- 0.3 Installation and setup of Linux using WSL
- 0.4 Basic Linux system orientation and navigation
- 0.5 Essential system preparation and command-line setup

◇ Module 1 – NGS Fundamentals & Linux Basics

This module covers the foundational concepts of NGS and introduces basic Linux commands required for handling sequencing data.

- 1.1 Introduction to Next-Generation Sequencing (NGS)
- 1.2 Overview of end-to-end NGS analysis workflow
- 1.3 Introduction to Linux shell environment
- 1.4 Hands-on practice with essential Linux commands
- 1.5 Understanding Linux file system structure
- 1.6 Setting up a mini-project workflow

◇ Module 2 – File Formats & Real NCBI Data

Participants will learn about biological data types, common file formats, and how to access and interpret real genomic data from public databases.

- 2.1 Types of biological data in genomics
- 2.2 Core file formats: FASTA, FASTQ, BAM, and VCF
- 2.3 Overview of public databases (NCBI, etc.)
- 2.4 Downloading reference genome (e.g., *E. coli*)
- 2.5 Understanding and interpreting FASTA file structure

◇ Module 3 – BLAST Analysis

This module introduces sequence similarity search using BLAST and compares local and web-based approaches.

- 3.1 Introduction to BLAST and its applications
- 3.2 Local BLAST vs. Web BLAST comparison
- 3.3 Hands-on execution of BLAST commands
- 3.4 Interpretation of BLAST output results
- 3.5 Comparative analysis using web-based BLAST tools

◇ Module 4 – Quality Control (QC)

Participants will learn how to assess and improve the quality of raw sequencing data using standard QC tools.

- 4.1 Importance of quality control in NGS
- 4.2 Understanding FASTQ format and Phred quality scores
- 4.3 Quality assessment of raw sequencing data
- 4.4 Read trimming and cleaning techniques
- 4.5 Post-processing quality evaluation
- 4.6 Interpretation of FastQC reports and modules

◇ Module 5 – Genome Assembly

This module focuses on assembling sequencing reads into complete genomes and evaluating assembly quality.

- 5.1 Concept of genome assembly
- 5.2 Applications and use-cases of assembly
- 5.3 Hands-on assembly using SPAdes
- 5.4 Validation and assessment of assembly results
- 5.5 Use of external tools for assembly confirmation

◇ Module 6 – Sequence Alignment

Participants will learn how sequencing reads are aligned to reference genomes and how alignment data is processed.

- 6.1 Introduction to sequence alignment

- 6.2 Workflow using BWA and SAMtools
- 6.3 Understanding SAM and BAM file formats
- 6.4 Evaluation of alignment metrics
- 6.5 Overview of alternative alignment tools

◇ Module 7 – Variant Calling (Conceptual)

This module introduces the principles of identifying genetic variations from sequencing data.

- 7.1 Definition and types of genetic variants
- 7.2 Basic logic of variant calling pipelines
- 7.3 Structure and interpretation of VCF files
- 7.4 Overview of industry-standard variant calling pipelines

Outcome of Section

By the end of this section, participants will be able to :

- Work efficiently in a Linux-based bioinformatics environment
- Handle real-world NGS datasets
- Perform key steps such as QC, assembly, alignment, and basic analysis
- Understand the complete NGS workflow from raw data to biological insights

Ready to Build Real-World AI Solutions in Biology?

Join the next generation of researchers combining Biology + AI + Innovation.

APPLY NOW

Visit Website : www.omniedgesci.com