Bioinformatics Scripting and Databases with Python/R Training Course

Professional Development Training Course Outline

Category	Biotechnology and Pharmaceutical Development	**Duration**	10 Days
Base Fee	$4,000 USD	**Delivery Mode**	Online / On-site Hybrid

Course Introduction

Bioinformatics Scripting and Databases with Python/R Training Course is engineered to bridge the skills gap between life sciences and data science, equipping researchers and analysts with the essential computational biology tools. Participants will master Python and R scripting for automated, efficient, and reproducible analysis of vast genomic data, proteomic data, and other high-throughput biological datasets. The curriculum emphasizes practical, hands-on training in managing, querying, and integrating data from various biological databases using industry-standard packages, ensuring graduates are immediately productive in demanding next-generation sequencing (NGS) analysis and precision medicine environments.

The program moves beyond theoretical concepts, focusing on the practical implementation of data manipulation, statistical analysis, and data visualization techniques critical for modern systems biology and drug discovery. By integrating the power of command-line tools with the advanced capabilities of Python's BioPython library and R's Bioconductor ecosystem, this course delivers a robust foundation in building professional-grade, reproducible bioinformatics pipelines. Graduates will gain the in-demand skills to independently design, execute, and troubleshoot complex computational genomics workflows, driving innovation and rigorous scientific inquiry in both academic and industrial settings.

Programme Curriculum

Bioinformatics Scripting and Databases with Python/R Training Course

Introduction

Bioinformatics Scripting and Databases with Python/R Training Course is engineered to bridge the skills gap between life sciences and data science, equipping researchers and analysts with the essential computational

biology tools. Participants will master Python and R scripting for automated, efficient, and reproducible analysis of vast genomic data, proteomic data, and other high-throughput biological datasets. The curriculum emphasizes practical, hands-on training in managing, querying, and integrating data from various biological databases using industry-standard packages, ensuring graduates are immediately productive in demanding next-generation sequencing (NGS) analysis and precision medicine environments.

The program moves beyond theoretical concepts, focusing on the practical implementation of data manipulation, statistical analysis, and data visualization techniques critical for modern systems biology and drug discovery. By integrating the power of command-line tools with the advanced capabilities of Python's BioPython library and R's Bioconductor ecosystem, this course delivers a robust foundation in building professional-grade, reproducible bioinformatics pipelines. Graduates will gain the in-demand skills to independently design, execute, and troubleshoot complex computational genomics workflows, driving innovation and rigorous scientific inquiry in both academic and industrial settings.

Course Duration

10 days

Course Objectives

1. Achieve fluency in Python/R for automated data processing and custom bioinformatics pipeline development.
2. Effectively query, retrieve, and parse data from major biological databases using programmatic interfaces.
3. Attain competence in Linux/Unix Shell scripting for efficient file manipulation and orchestrating workflows.
4. Implement best practices for handling, quality control (QC), and preprocessing Next-Generation Sequencing (NGS) raw data.
5. Script fundamental sequence analysis tasks, including sequence alignment and motif finding.
6. Apply R and Bioconductor for differential gene expression (DGE) analysis of RNA-seq and other -omics data.
7. Design and implement relational databases (SQL) for storing and querying complex genomic annotations.
8. Develop and document robust, scalable workflows using version control and workflow management systems
9. Create publication-quality, interactive visualizations of genomic data and statistical results using Python/R libraries
10. Apply introductory machine learning algorithms to biomedical data for predictive modeling.
11. Use scripting to analyze protein sequence and 3D structure data and interact with PDB and related resources.
12. Perform pathway and network analysis to interpret large-scale gene expression and proteomic results.
13. Acquire essential skills for debugging scripts and optimizing code performance for big data bioinformatics challenges.

Target Audience

1. Life Scientists/Biologists.
2. Bioinformatics Analysts
3. Data Scientists
4. M.Sc./Ph.D. Students.

5. Clinical Researchers
6. Pharmaceutical/Biotech R&D.
7. Software Developers.
8. Core Facility Staff.

Course Modules

Module 1: Introduction to Bioinformatics and Scripting Fundamentals

- Overview of Genomics, Transcriptomics, Proteomics, and Metabolomics.
- Installing Python, R, RStudio, and essential Linux command-line tools.
- Variables, Data Types, Control Structures in Python/R.
- Reading and writing standard FASTA, FASTQ, and CSV file formats.
- Case Study: Writing a Python script to count nucleotide and amino acid frequencies from a FASTA file.

Module 2: Linux Shell Scripting for Bioinformatics Workflows

- Navigation File Management Permissions
- Using powerful command-line utilities like grep, awk, sed, and cut for quick data parsing.
- Building simple, chained workflows using standard input/output.
- Writing reusable .sh scripts for batch job automation.
- Case Study: Developing a Bash script to automate the QC reporting for a directory of NGS samples.

Module 3: Python Core Data Structures and Functions

- Python Data Structures.
- Modular Programming.
- Object-Oriented Programming (OOP) Basics.
- Using try-except blocks for robust script execution.
- Case Study: Creating a dictionary to store codon usage frequency and a function to translate a DNA sequence.

Module 4: R Core Concepts and Data Manipulation

- R Data Structures.
- Introduction to the Tidyverse ecosystem
- Filtering, arranging, summarizing, and mutating data frames using dplyr.
- Efficiently loading and saving various data formats in R.
- Case Study: Cleaning and restructuring a messy clinical metadata table in R for downstream statistical analysis.

Module 5: BioPython for Sequence Analysis

- Creating and manipulating sequence objects and records.
- Reading and writing complex biological file formats
- Sequence Alignment.
- Translating and Transcription.
- Case Study: Parsing a GenBank file to extract genes, features, and annotated metadata into a structured report.

Module 6: Bioconductor for Genomic Data

- Bioconductor Ecosystem.
- Core Data Structures.
- Installation and Management.
- Annotation Data
- Case Study: Loading a public RNA-seq count matrix into a SummarizedExperiment object for preliminary inspection.

Module 7: Interacting with Biological Databases (NCBI, EBI)

- Programmatic searching and data retrieval using Bio.Entrez (Python) and rentrez (R).
- Submitting, retrieving, and parsing BLAST search results.
- Interacting with RESTful APIs from databases like UniProt and Ensembl.
- Strategies for efficient and compliant data retrieval to avoid overwhelming public servers.
- Case Study: Scripting the download of all human RefSeq protein sequences associated with a specific gene list.

Module 8: Introduction to Relational Databases (SQL)

- SQL Fundamentals
- Database Design.
- Python/R and SQL
- Data Integration
- Case Study: Designing an SQL database schema to store experimental results and linking it to a table of Gene Ontology (GO) annotations.

Module 9: Next-Generation Sequencing (NGS) Data Processing

- Introduction to aligners and key file formats
- Implementing Python/R scripts for post-alignment quality filtering.
- Overview of SNP/Indel detection and the VCF file format.
- Combining Linux and Python scripts to create an end-to-end NGS processing mini-pipeline.
- Case Study: Using pysam to read a BAM file, count reads mapping to a specific genomic region, and calculate coverage.

Module 10: RNA-Seq Differential Gene Expression (DGE) Analysis

- Understanding statistical models for count data
- Normalization and filtering of count matrices in R.
- Bioconductor Packages.
- Generating volcano plots, MA plots, and identifying significantly differentially expressed genes.
- Case Study: Executing a complete DESeq2 pipeline on a published RNA-seq dataset comparing two biological conditions.

Module 11: Advanced Data Visualization

- Python Visualization.
- Mastering ggplot2 for layered, aesthetic, and publication-quality figures.
- Genomic Visualization.
- Introduction to tools like Plotly or Shiny for dynamic data exploration.
- Case Study: Generating a high-resolution Volcano Plot of DGE results and a heat map of the top 50 differentially expressed genes.

Module 12: Machine Learning for Biomedical Data

- ML Fundamentals.
- Applying K-means or Hierarchical Clustering to gene expression data
- Building simple models to predict disease state from molecular profiles.
- Using PCA and t-SNE to visualize high-dimensional data.
- Case Study: Training a classification model in Python to distinguish between healthy and diseased samples based on metabolomic profiles.

Module 13: Reproducible and Scalable Workflows

- Using Git and GitHub for collaborative code development and tracking changes.
- Introduction to Docker for creating reproducible execution environments.
- Workflow Management Systems.
- Writing modular, well-documented, and tested code.
- Case Study: Converting a collection of Bash and Python scripts into a single, defined, and version-controlled Snakemake pipeline.

Module 14: Structural and Proteomics Bioinformatics

- Protein Data Bank (PDB).
- Structural Analysis
- Mass Spectrometry Data.
- Mapping protein sequences to functional domains
- Case Study: Scripting the calculation of the Root Mean Square Deviation (RMSD) between two conformations of a protein structure.

Module 15: Pathway and Network Analysis

- Performing Gene Ontology (GO) and KEGG pathway enrichment analysis.
- Network Biology
- Using R/Python to visualize and analyze networks
- Linking DGE results back to biological pathways for mechanistic hypothesis generation.
- Case Study: Applying GO enrichment to the significantly differentially expressed genes from the RNA-seq case study to identify affected biological processes.

Training Methodology

The course employs an intensive Blended Learning approach, combining Interactive Lectures with a strong emphasis on Hands-on Labs and Real-World Case Studies.

- Project-Based Learning.
- Live Coding Demonstrations.
- Pair Programming/Collaborative Work.
- Cloud/Containerized Environment.

Register as a group from 3 participants for a Discount

Send us an email: info@fineskilltrainingcenter.org or call +254769199797

Certification

Upon successful completion of this training, participants will be issued with a globally- recognized certificate.

Tailor-Made Course

We also offer tailor-made courses based on your needs.

Key Notes

a. The participant must be conversant with English.

b. Upon completion of training the participant will be issued with an Authorized Training Certificate

c. Course duration is flexible and the contents can be modified to fit any number of days.

d. The course fee includes facilitation training materials, 2 coffee breaks, buffet lunch and A Certificate upon successful completion of Training.

e. One-year post-training support Consultation and Coaching provided after the course.

f. Payment should be done at least a week before commence of the training, to Fineskill Training Center account, as indicated in the invoice so as to enable us prepare better for you.

Upcoming Scheduled Sessions

Start Date	End Date	Location	Price
23 Nov 2026	04 Dec 2026	Online	$2,000
23 Nov 2026	04 Dec 2026	Physical	$3,000
23 Nov 2026	04 Dec 2026	Physical	$6,000
23 Nov 2026	04 Dec 2026	Physical	$6,000
23 Nov 2026	04 Dec 2026	Physical	$12,000
23 Nov 2026	04 Dec 2026	Physical	$9,000
23 Nov 2026	04 Dec 2026	Physical	$12,000
23 Nov 2026	04 Dec 2026	Physical	$12,000

Ready to enroll or request a customized program? Book online or contact us:

Register Online Now

Email: info@fineskilltrainingcenter.com | Website: www.fineskilltrainingcenter.com