Biostatistics for Next-Generation Sequencing (NGS) 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

Next-Generation Sequencing (NGS) has irrevocably transformed modern genomics and molecular biology, generating massive, complex datasets that require sophisticated quantitative analysis. Biostatistics for Next-Generation Sequencing (NGS) Training Course is meticulously designed to bridge the gap between raw sequencing data and biological knowledge. Participants will gain critical bioinformatics skills and a deep understanding of the statistical foundations necessary for robust experimental design, rigorous quality control, and accurate interpretation of high-throughput genomic data. Mastery of these quantitative techniques is paramount for advancing research in precision medicine, cancer genomics, infectious disease surveillance, and functional genomics.

This program goes beyond simple tool tutorials, focusing on the quantitative principles that underpin all major NGS applications, including RNA-Seq, Whole-Exome Sequencing (WES), and single-cell genomics. We will equip you with hands-on proficiency in industry-standard computational environments like R and Python, and key bioinformatics tools. By completing this training, you will be prepared to independently design, execute, and interpret reproducible NGS data analysis workflows, accelerating your organization's research productivity and driving data-driven discovery in the rapidly evolving landscape of genomic data science.

Programme Curriculum

Biostatistics for Next-Generation Sequencing (NGS) Training Course

Introduction

Next-Generation Sequencing (NGS) has irrevocably transformed modern genomics and molecular biology, generating massive, complex datasets that require sophisticated quantitative analysis. Biostatistics for Next-Generation Sequencing (NGS) Training Course is meticulously designed to bridge the gap between raw sequencing data and biological knowledge. Participants will gain critical bioinformatics skills and a deep understanding of the statistical foundations necessary for robust experimental design, rigorous quality control, and accurate interpretation of high-throughput genomic data. Mastery of these quantitative techniques is paramount for advancing research in precision medicine, cancer genomics, infectious disease surveillance, and functional genomics.

This program goes beyond simple tool tutorials, focusing on the quantitative principles that underpin all major NGS applications, including RNA-Seq, Whole-Exome Sequencing (WES), and single-cell genomics. We will equip you with hands-on proficiency in industry-standard computational environments like R and Python, and key bioinformatics tools. By completing this training, you will be prepared to independently design, execute, and interpret reproducible NGS data analysis workflows, accelerating your organization's research productivity and driving data-driven discovery in the rapidly evolving landscape of genomic data science.

Course Duration

10 days

Course Objectives

1. Master statistical principles for NGS experimental design
2. Execute command-line operations and Linux fundamentals for high-throughput data processing.
3. Perform rigorous quality control (QC) and data preprocessing on raw sequencing reads
4. Implement read alignment and variant calling pipelines for DNA-Seq data.
5. Conduct robust differential gene expression (DGE) analysis for RNA-Seq data using Bioconductor packages
6. Analyze complex genomic variations, including SNPs, Indels, and Copy Number Variants
7. Apply biostatistical methods for analyzing single-cell RNA-Seq data
8. Evaluate and interpret data from ChIP-Seq and ATAC-Seq for epigenomics studies.
9. Develop and run pipelines for metagenomic sequencing and microbial community profiling
10. Utilize advanced machine learning techniques for genomic data classification and feature selection.
11. Conduct functional and pathway enrichment analysis to derive biological insights.
12. Ensure data reproducibility and implement FAIR data principles in genomic workflows.
13. Communicate and visualize complex genomic and biostatistical results effectively for translational research.

Target Audience

1. **Biologists/Molecular Scientists:** Seeking to analyze their own NGS data.
2. **Bioinformaticians:** Wishing to deepen their statistical and advanced application skills (e.g., scRNA-Seq, metagenomics).
3. **Clinicians/Pathologists:** Involved in **clinical genomics** and interpreting sequencing results for diagnosis.
4. **Research Associates & Technicians:** Responsible for running and analyzing sequencing experiments.
5. **Data Scientists/Statisticians:** Transitioning into the **genomic data science** domain.
6. **Graduate Students & Post-Docs:** Specializing in genomics, genetics, and computational biology.
7. **Pharmaceutical/Biotech R&D Staff:** Working on drug target identification and biomarker discovery.
8. **Public Health Officials:** Focusing on pathogen surveillance and population **genomics**.

Course Modules

Module 1: Foundational Biostatistics & NGS Overview

- Hypothesis testing, probability, type I/II errors, and p-value correction
- Illumina, PacBio, Oxford Nanopore platforms, read lengths, and applications.
- FASTQ, SAM/BAM, VCF, and BED files.
- Essential navigation and scripting for pipeline automation.
- Case Study: Analyzing the impact of sequencing depth on variant discovery sensitivity.

Module 2: Quality Control (QC) and Read Preprocessing

- Using FastQC to evaluate read quality metrics.
- Implementing Trimmomatic and Cutadapt.
- Identifying and mitigating bias
- Generating aggregate reports for project-wide quality assurance.
- Case Study: Trimming a public cancer dataset to improve alignment efficiency and accuracy.

Module 3: Reference Genomes and Sequence Alignment

- Preparing reference genomes with tools like Bowtie2 and BWA.
- Understanding local and global alignment and gap penalties.
- Practical application of BWA-MEM for short-read alignment.
- Sorting, indexing, and visualization using SAMtools and IGV.
- Case Study: Mapping Whole Genome Sequencing reads from a patient to the human reference genome

Module 4: Variant Calling Pipelines (DNA-Seq)

- Key differences in methodology and statistical models.
- Step-by-step workflow for realignment and base quality recalibration.
- Using GATK HaplotypeCaller and FreeBayes.
- Applying hard filters and Variant Quality Score Recalibration
- Case Study: Identifying a pathogenic germline SNP in a Familial Hypercholesterolemia WES dataset.

Module 5: Variant Annotation and Interpretation

- Functional Annotation
- Filtering by population frequency and predicted effect
- Querying ClinVar and CIViC for clinical relevance.
- Introduction to regulatory region analysis.
- Case Study: Prioritizing a single likely causal somatic mutation in a tumor exome for targeted therapy.

Module 6: RNA-Seq Data Analysis Fundamentals

- RNA-Seq Applications
- Alignment-based and pseudo-alignment methods.
- Count Matrix Generation.
- TMM, RPKM, FPKM, and TPM to account for library size and gene length.
- Case Study: Quantifying transcript abundance for over 20,000 genes in a disease vs. control cohort.

Module 7: Differential Gene Expression (DGE) with R

- Applying the Negative Binomial Distribution for count data.

- Hands-on DGE using DESeq2 and edgeR.
- Correcting for multiple testing to control False Discovery Rate (FDR).
- Log-Fold Change (LFC) Shrinkage.
- Case Study: Identifying significant differentially expressed genes (DEGs) in a drug-treated cell line.

Module 8: Advanced RNA-Seq Topics

- Batch Effect Correction.
- Modeling gene expression changes over multiple time points.
- Introduction to tools like MISO and rMATS.
- Gene Set and Pathway Analysis
- Case Study: Correcting a major batch effect in a multi-center gene expression study to reveal true biological signal.

Module 9: Single-Cell RNA-Seq (scRNA-Seq) Introduction

- scRNA-Seq Technologies.
- Quality metrics, filtering, and normalization of UMI count matrices.
- Applying PCA and t-SNE/UMAP for data visualization.
- Identifying distinct cell populations using Seurat or Scanpy.
- Case Study: Identifying novel immune cell sub-types in a tumor microenvironment scRNA-Seq dataset.

Module 10: Metagenomics and Microbiome Analysis

- Sequencing Strategies.
- Classifying reads using tools like Kraken2 and MetaPhlAn.
- Statistical measures for microbial community comparison.
- Inferring metabolic pathways (HUMAnN).
- Case Study: Comparing the gut microbiome profiles of lean vs. obese individuals and identifying associated microbial species.

Module 11: Epigenomics and Functional Genomics

- Peak calling for transcription factor binding sites
- Identifying regions of chromatin accessibility.
- Biostatistical analysis of enriched regions.
- Linking epigenetic marks to gene expression changes.
- Case Study: Identifying novel regulatory elements activated by a specific transcription factor in response to stress.

Module 12: Advanced Statistical Modeling

- Kaplan-Meier curves and Cox Proportional Hazards model in cancer genomics.
- Methods for detecting large-scale genomic alterations.
- Introduction to classification and regression for genomic biomarker discovery.
- Cross-validation and performance metrics
- Case Study: Building an ML model to predict patient survival based on a set of core differentially expressed genes.

Module 13: Data Visualization and Reporting

- Creating publication-quality plots using ggplot2 in R.

- Using tools like Plotly and Shiny for dynamic data exploration.
- Generating reproducible reports with R Markdown/Jupyter Notebooks.
- Best practices for robust and transparent workflows.
- Case Study: Designing a comprehensive 'data package' report for a large-scale exome sequencing project.

Module 14: Practical R and Python for Genomics

- Data structures, control flow, and functions.
- Efficient data manipulation with dplyr and Tidyverse.
- Introduction to Pandas and NumPy for genomic data handling.
- Exploring key packages for genomic analysis.
- Case Study: Writing a modular R script to automate the batch processing of FastQC reports.

Module 15: Reproducibility and Career Skills

- Making data Findable, Accessible, Interoperable, and Reusable.
- Introduction to Docker and Singularity for environment consistency.
- Using Git for collaborative coding and pipeline tracking.
- Overview of resources for large-scale genomic analysis (AWS, Google Cloud).
- Case Study: Creating a reproducible Docker image containing a fully functional RNA-Seq analysis pipeline.

Training Methodology

The course adopts a highly practical, Blended Learning approach to ensure deep understanding and immediate applicability:

1. Interactive Lectures.
2. Hands-on Labs.
3. Case Study-Based Learning.
4. Group Discussions & Peer Learning.
5. Capstone Project.

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