

100%  
Placement  
Assistance



# AI & MACHINE LEARNING IN CANCER GENOMICS



**HANDS-ON TRAINING,  
INDEPENDENT RESEARCH  
PROJECTS & PAPER  
PUBLICATION GUIDANCE**

100% Placement Assistance



A comprehensive research-oriented program at the intersection of genomics, machine learning, and precision oncology

**EXCLUSIVE PROGRAM**  
**CALL TOLL FREE: 1800-1200-1818**

Website

[stores.biotecnika.org](http://stores.biotecnika.org)

# INTRODUCTION

## Why AI in Cancer Genomics?



Cancer is fundamentally a genomic disease. Advances in next-generation sequencing have enabled researchers to generate massive volumes of genomic data describing mutations, copy number changes, gene expression alterations, and tumor evolution patterns across thousands of patients.

However, raw genomic data alone does not lead to insights. To extract clinically and biologically meaningful information, modern cancer research increasingly relies on:

- Machine Learning (ML) for pattern discovery and prediction
- Deep Learning (DL) for sequence modeling and complex feature extraction
- Multi-omics integration to understand cancer as a system
- Explainable AI to ensure clinical trust and regulatory compliance

This program is designed to train participants in how AI is actually used in cancer genomics research today, not just conceptually, but practically and rigorously.



Page 01



# Program Philosophy & Learning Approach

This is not a short skills course and not a tool-demo workshop. The AI & Machine Learning in Cancer Genomics program is designed as a research training experience, closely resembling how work is carried out in:

- Academic cancer genomics labs
- Computational oncology research group
- Precision medicine and translational research teams

## CORE PRINCIPLES OF THE PROGRAM



# Who This Program Is Designed For?

## ELIGIBILITY

This program is intended for learners who want to go beyond surface-level bioinformatics and build research depth.

### Ideal participants include:

- BSc / MSc students in Life Sciences, Biotechnology, Genetics, Bioinformatics, Pharmacy
- PhD aspirants preparing for computational or cancer genomics research
- PhD scholars seeking advanced AI-based genomics exposure
- Working professionals in pharma, CROs, diagnostics, genomics startups
- Biologists transitioning into AI & data-driven research
- Data science professionals interested in healthcare and genomics

## BACKGROUND NEEDED



Basic understanding  
of biology (genes,  
DNA, mutations)



No prior ML or DL  
expertise required —  
foundations are covered



# Program Outcomes

By completing this program, participants will be able to:

- Understand cancer genomics data structures and biological meaning
- Build end-to-end genomics analysis pipelines
- Apply machine learning models to cancer classification and survival prediction
- Develop deep learning models for genomic sequence analysis
- Perform multi-omics data integration
- Analyze tumor heterogeneity, evolution, and microenvironment
- Design AI-driven precision oncology pipelines
- Interpret results in a biologically and clinically meaningful way
- Execute an independent research project
- Prepare a manuscript suitable for journal or conference submission



**Machine & deep learning for cancer genomics**



**Multi-omics integration & tumor biology insights**



**Precision oncology pipeline development**



# Detailed Training Curriculum

Modules	Session Topic	Subtopics
Module 1 – Foundations  (Genomics + Python + ML Basics)	Intro to Cancer Genomics	What is cancer genomics? Types of genomic alterations: SNVs, INDELs, CNVs, SVs Tumor heterogeneity, clonal evolution File formats: FASTQ, BAM, VCF, GTF, BED
	Tools overview	GATK, bcftools, samtools Xenabrowser, cBioPortal
	Python Foundations for Bioinformatics	Numpy, Pandas, Matplotlib Biopython basics Reading FASTA/VCF files
	Complete Genomics end-to-end pipeline	FASTQ → BAM → VCF → annotation to give practical cancer genomics workflow experience
	Essential Statistics for Genomics	Hypothesis testing: t-test, ANOVA p-value, FDR, Benjamini–Hochberg Distribution of biological data
	Machine Learning Essentials	Supervised vs. unsupervised learning Regression & classification Train/val/test splits, Build a simple logistic regression classifier.
	Feature Engineering in Genomics	Encoding sequences k-mer generation One-hot encoding Mutation matrices
	Exploratory Data Analysis (EDA)	Plot mutation frequency Volcano plot basics Correlation heatmaps
	Module 1 Practical	Explore TCGA/ICGC dataset portals - Integrate real TCGA/ICGC datasets from Week 1 for actual practical learning with actual cancer genomics data used in research.

QnA / Open Discussion Session



# Detailed Training Curriculum

Modules	Session Topic	Subtopics
Module 2 – ML for Genomics (Intermediate)	Classical ML for Genomics	SVM, Random Forest, XGBoost Model evaluation (ROC-AUC, F1, Recall)
	Dimensionality Reduction	<b>Practical:</b> PCA, t-SNE and UMAP, Feature selection
	Clustering Cancer Samples	<b>Practical:</b> K-means, Hierarchical clustering, Consensus clustering
	Survival Analysis + ML	<b>Practical:</b> Kaplan-Meier, Cox regression, ML integration for survival prediction
	Mutational Signatures (COSMIC)	SBS, DBS, ID signatures SigProfiler usage Practical: Extract mutational signatures from sample data.
	Genomic Data Integration	Gene expression + mutation + CNV Multi-omics ML models data analysis and visualization using Databases and tools
	<b>Module 2 Practical</b>	Build a cancer subtype classifier (XGBoost) using multi-omics TCGA data.
	Classical ML for Genomics	SVM, Random Forest, XGBoost Model evaluation (ROC-AUC, F1, Recall)
QnA / Open Discussion Session		



# Detailed Training Curriculum

QnA / Open Discussion Session		
Modules	Session Topic	Subtopics
<b>Module 3 – Deep Learning for Cancer Genomics</b>	Deep Learning Basics (PyTorch/TensorFlow)	Neural networks Loss functions Optimization
	CNNs for Genomic Sequences	Sequence convolution Motif detection
	RNNs & LSTMs	Sequence classification Mutation prediction: LSTM on DNA sequences.
	Transformers in Genomics	Attention mechanisms GenomicBERT, DNABERT
	Variational Autoencoders (VAE)	Latent space Tumor heterogeneity representation
	Graph Neural Networks (GNNs) for Cancer	Gene networks PPI graphs Pathway-level cancer prediction
	<b>Module 3 Practical</b>	Transformer or GNN model to classify cancer vs normal based on mutations.

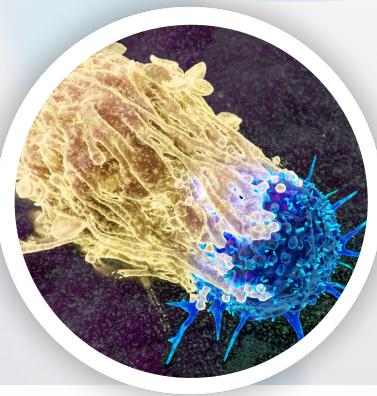


# Detailed Training Curriculum

QnA / Open Discussion Session		
Modules	Session Topic	Subtopics
Module 4 – Advanced Cancer Genomics + Clinical AI	Driver Mutation Prediction	Tools: MutSigCV, OncoKB ML for identifying driver vs passenger, gene classifier.
	Tumor Microenvironment + ML	Immune cell deconvolution
	Single-Cell Cancer Genomics + ML	scRNA-seq workflow Clustering + trajectory inference ML for tumor purity & clones Practical: Classify malignant vs non-malignant cells.
	Drug Response Prediction	GDSC, CCLE datasets ML for IC50 prediction Practical: Train model to predict drug sensitivity/toxicity.
	Precision Oncology Pipelines	Variant annotation (VEP, Clinvar, etc) Clinical actionability ML-based treatment recommendation
	AI in Clinical Diagnostics	Liquid biopsy ML models Biomarker prediction Radiogenomics overview Practical: Integrate mutation + imaging features
	Ethical AI + Regulatory Standards	Explainable AI (SHAP, LIME) Bias in clinical ML FDA/ICMR guidelines
Capstone Project discussion		Project topic discussion, Test and project next steps
QnA / Open Discussion Session		



# Project Specializations AI & Machine Learning in Cancer Genomics



Projects are segregated by core research specialization i.e. what scientific cancer problem the project focuses on.

1

## Cancer Classification & Subtyping

**(Projects focused on identifying cancer types, subtypes, and sample grouping)**

- Machine Learning-Based Cancer Subtype Classification Using Somatic Mutation Profiles – 3 months
- Multi-Modal Cancer Subtype Discovery Using Integrated Genomic Features – 6 months
- Cancer Type Classification Using Gene Expression Data – 3 months
- Dimensionality Reduction and Clustering of Cancer Genomic Samples – 3 months
- Deep Learning-Based Classification of Cancer vs Normal Using Mutation Encodings – 3 months



# Project Topics CONTD....



**2**

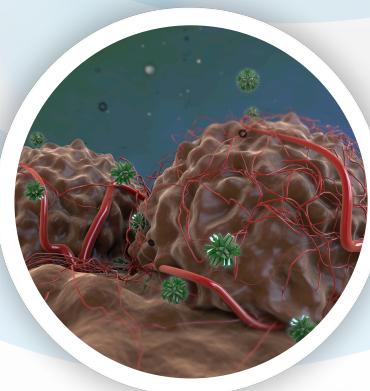
## Mutation Impact, Drivers & Variant Interpretation

**(Projects focused on mutation significance, pathogenicity, and functional interpretation)**

- Driver vs Passenger Mutation Classification Using Supervised Learning – 3 months
- Explainable AI for Identifying Driver Mutations from Cancer Variant Data – 3 months
- Mutation type, Somatic Variant Impact Prediction Using ML – 6 months
- Pharmacogenomic Variant Identification in CYP2D6 and Cancer Therapy – 6 months
- End-to-End AI Pipeline for Somatic Variant Interpretation in Precision Oncology – 12 months
- AI-Driven Identification of Novel Cancer Driver Pathways – 12 months
- Basic Mutation Frequency Analysis and Visualization in TCGA Breast Cancer Data – 3 months
- k-mer Based Classification of COSMIC Mutational Signature Types – 3 months
- Support Vector Machine Prediction of COSMIC Signature Contributions in Pan-Cancer Samples – 6 months



# Project Topics CONTD....



**3**

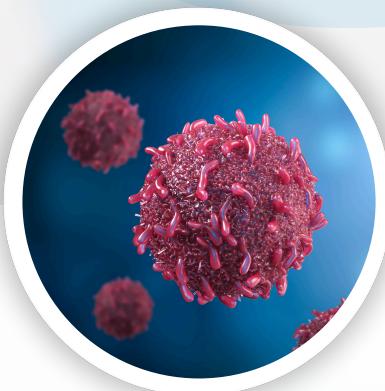
Tumor Evolution, Heterogeneity & Single-Cell Genomics

**(Projects studying intra-tumor diversity  
clonal evolution, and cellular composition**

- Analysis of Tumor Microenvironment Using Single-Cell Genomics – 6 months
- Deep Generative Modeling of Tumor Heterogeneity Using Variational Autoencoders – 12 months
- AI-Driven Detection of Tumor Evolution and Clonal Heterogeneity Using Longitudinal Cancer Genomics – 12 months



# Project Topics CONTD....



4

Survival, Prognosis & Risk Stratification

**(Projects focused on patient outcomes and disease progression prediction)**

- Survival Outcome Stratification Using Cox Regression and ML Models – 3 months
- Prediction of Cancer Progression Using Multi-Omics Data – 6 months
- AI-Driven Multi-Omics Modeling of Cancer Progression and Patient Survival – 12 months
- Radiogenomic Feature Fusion for Cancer Risk Stratification – 12 months
- t-SNE and UMAP Visualization of Copy Number Variations Across Cancer Types – 6 months



# Project Topics CONTD....



5

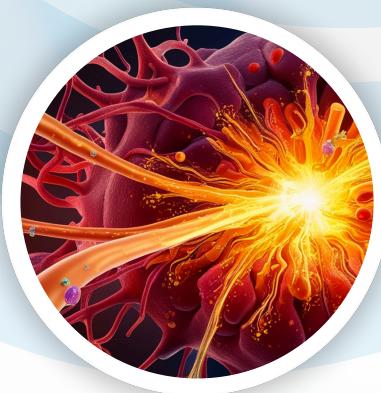
Drug Response, Therapy & Precision Oncology

**(Projects linking genomics with treatment response and therapeutic decision-making)**

- Drug Sensitivity Prediction Using Cancer Cell Line Genomics – 6 months
- Predicting Drug Response in Cancer Using ML – 6 months
- AI-Based Multi-Omics Fusion for Predicting Cancer Therapy Response – 6 months
- Potential therapeutic interventions against any type of cancer via the CADD approach – 6 months
- Genomic variance effector prediction and mutant protein modelling for any type of cancer – 6 months



# Project Topics CONTD....



6

Advanced AI Architectures in  
Cancer Genomics

**(Projects emphasizing modern AI model architectures rather than a single biological task)**

- Transformer-Based Mutation Pattern Learning for Cancer Classification – 6 months
- Graph Neural Network Modeling of Gene Interaction Networks in Cancer – 6 months
- Graph Neural Network (GNN)-Based Cancer Genome Modeling for Precision Oncology – 12 months
- Graph-Based Pathway Enrichment Analysis for Predicting Therapeutic Targets in Oncology – 12 months



# Project Topics CONTD....



7

Clinical AI, Explainability & Regulatory Systems

**(Projects designed for clinical deployment, trust, and regulatory readiness)**

- Explainable AI for Clinical Variant Prioritization – 6 months
- Regulatory-Compliant Clinical AI Model for Genomic Diagnostics – 12 months



# Project Execution & Mentorship

**Each project follows a structured research workflow:**

1. Research question formulation
2. Literature review
3. Dataset selection
4. Data preprocessing & feature engineering
5. Model development
6. Evaluation & validation
7. Biological and clinical interpretation
8. Result visualization
9. Manuscript drafting



Structured, end-to-end research project execution



Dedicated mentorship for publication readiness



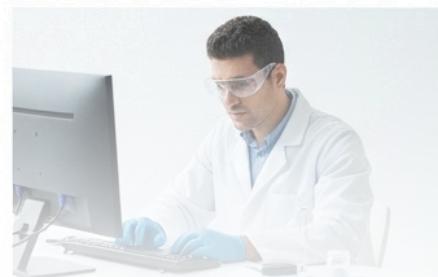
# Paper Publication Guidance

## Participants receive guidance on:

- Manuscript structure and scientific writing
- Figures and result presentation
- Journal or conference selection
- Ethical and reproducibility considerations

**Publication acceptance depends on peer review guidance is provided throughout.**

# Career Pathways After This Program



This program prepares learners for long-term careers, not just certificates.

## Research & Academic Careers

- PhD in Cancer Genomics / Computational Biology
- Research Fellow / Research Assistant
- Academic lab-based cancer genomics research

## Industry & Applied Research Roles

- Cancer Bioinformatician
- Computational Biologist
- Genomics Data Scientist
- Precision Oncology Analyst
- Translational Research Scientist



# Career Pathways After This Program



## AI & Healthcare Technology Roles

- AI Research Associate (Healthcare)
- Clinical AI Model Developer
- Multi-omics Data Scientist
- Explainable AI Specialist

## Pharma, CRO & Diagnostics

- Drug response & biomarker analysis roles
- Clinical genomics interpretation teams
- Regulatory-compliant AI development roles





**This program is best suited for learners who are serious about research depth, AI-driven cancer genomics, and long-term academic or industry careers.**

*Build research depth.  
Work with real cancer data.  
Create publishable outcomes.*



**Email**  
[info@biotecnika.org](mailto:info@biotecnika.org)



**Website**  
[stores.biotecnika.org](http://stores.biotecnika.org)



**Phone**  
1800-1200-1818

