

MODULE 2

Advanced Transcriptomics & AI for Precision Biology

Single-Cell RNA-Seq • Machine Learning • Disease Prediction

Explore the frontier of modern bioinformatics by learning how transcriptomic data is transformed into biological discoveries and predictive disease models. This module combines **single-cell transcriptomics** and **machine learning**, providing hands-on experience with real-world datasets and industry-standard tools.

Module Duration

31 July 2026 – 13 August 2026

Class Timings: 10:00 AM – 4:00PM

Week 1 Classes (Single-Cell RNA-Seq)

31 July 2026

01 August 2026

02 August 2026

Independent Assignment Period

03 August 2026 – 06 August 2026

Participants will independently analyze a complete scRNA-seq dataset and prepare a research-style report and presentation.

Week 2 Classes (Machine Learning for Transcriptomics)

07 August 2026

08 August 2026

09 August 2026

Capstone Project Period

10 August 2026 – 13 August 2026

Participants will develop and evaluate disease-classification models using transcriptomic datasets and present their findings.

Total Learning Commitment

6 Live Interactive Classes

36 Hours of Instructor-Led Training

2 Independent Research Projects

Real Biological Datasets Research-Oriented

Assignments Project Presentation & Evaluation

Week 1: Single-Cell RNA Sequencing (scRNA-seq)

Theme:

From Raw Sequencing Data to Cellular Insights

Participants will learn the complete workflow of single-cell RNA sequencing analysis using **Cell Ranger** and **Scanpy**, from raw sequencing reads to biological interpretation.

Class 1

Foundations of Transcriptomics & Single-Cell Biology

Topics

- Introduction to transcriptomics
- Bulk RNA-seq vs Single-Cell RNA-seq
- Applications of scRNA-seq in modern research
- Experimental workflow of single-cell sequencing
- Understanding sequencing outputs
- FASTQ files and sequencing reads
- Introduction to Cell Ranger
- Structure and interpretation of Cell Ranger outputs

Hands-on

- Exploring a real scRNA-seq dataset
- Understanding Cell Ranger-generated files
- Preparing datasets for downstream analysis

Class 2

Data Processing & Quality Control

Topics

- Introduction to Scanpy
- Building AnnData objects

- Quality control metrics
- Identifying low-quality cells
- Detecting doublets and outliers
- Normalization techniques
- Identification of highly variable genes

Hands-on

- Performing quality control
- Cell filtering strategies
- Data normalization
- Feature selection

Class 3

Clustering, Visualization & Biological Interpretation

Topics

- Principal Component Analysis (PCA)
- Neighborhood graph construction
- Clustering algorithms
- UMAP and t-SNE visualization
- Marker gene identification
- Differential expression analysis
- Cell type annotation
- Biological interpretation of results

Hands-on

- Identifying cellular subpopulations
- Discovering marker genes
- Annotating cell clusters
- Generating publication-quality figures

Week 1 Assignment & Project

Independent Single-Cell Research Project

Each participant will receive an unseen scRNA-seq dataset.

Tasks

1. Import and preprocess the dataset
2. Perform quality control
3. Normalize and identify highly variable genes
4. Generate PCA and UMAP visualizations
5. Perform clustering analysis

6. Identify marker genes for each cluster
7. Annotate putative cell types
8. Investigate biologically meaningful patterns

Deliverables

- Complete Scanpy analysis notebook
- UMAP visualizations
- Cluster annotation report
- Marker gene tables
- Final presentation summarizing biological findings

Expected Outcome

Participants will independently perform a complete single-cell transcriptomic analysis and communicate their findings in a research-oriented format.

Week 2: Machine Learning for Transcriptomics

Theme:

From Gene Expression to Disease Prediction

Participants will learn how machine learning models can be trained using transcriptomic datasets to classify biological phenotypes and identify predictive biomarkers.

Class 4

Preparing Biological Data for Machine Learning

Topics

- Introduction to machine learning in bioinformatics
- Types of learning algorithms
- Structure of gene expression matrices
- Biological metadata and labels
- Data cleaning and preprocessing
- Feature scaling and normalization
- Train-test splitting strategies
- Understanding model bias and variance

Hands-on

- Preparing bulk RNA-seq datasets
- Data preprocessing workflow
- Feature engineering

Class 5

Feature Selection & Model Development

Topics

- High-dimensional biological data challenges
- Feature selection strategies
- Statistical feature filtering
- Recursive feature elimination
- Random Forest feature importance
- Training classification models
- KNN
- Decision Trees
- Random Forest

Hands-on

- Selecting informative genes
- Training multiple classification models
- Comparing model performance

Class 6

Model Evaluation & Biological Interpretation

Topics

- Confusion matrix
- Accuracy, Precision, Recall
- F1 Score
- ROC and AUC analysis
- Cross-validation
- Model interpretation
- Biomarker identification
- Translating predictions into biological insights

Hands-on

- Model evaluation
- Performance comparison
- Biomarker discovery
- Result visualization

Week 2 Assignment & Capstone Project

Disease Classification Challenge

Each participant will receive an independent bulk RNA-seq gene expression dataset containing healthy and disease samples.

Tasks

1. Explore and understand the dataset
2. Preprocess gene expression data
3. Perform feature selection
4. Train at least three classification models
5. Evaluate model performance
6. Compare model predictions
7. Identify top predictive genes
8. Interpret biological significance

Deliverables

- Complete machine learning pipeline
- Trained classification models
- Performance evaluation report
- Biomarker candidate list
- Final project presentation

Expected Outcome

Participants will independently develop, evaluate, and interpret machine learning models capable of distinguishing biological phenotypes using transcriptomic data.

Module Outcomes

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

Analyze scRNA-seq datasets using Cell Ranger and Scanpy

Discover cellular heterogeneity and identify marker genes

Interpret transcriptomic datasets biologically

Apply machine learning to gene expression data

Build disease-classification models from RNA-seq datasets

Identify candidate biomarkers and predictive signatures

Present bioinformatics findings in a research and industry-ready format