COURSE TITLE	**R FOR BIOLOGISTS**
COURSE CODE	**01CB0608**
COURSE CREDITS	**2**

Objective:

1 The course aims to introduce students to the basics of R programming and RStudio for biological data analysis.

2 To develop proficiency in using R for data handling, statistical analysis, and interpretation of biological datasets.

Course Outcomes: After completion of this course, student will be able to:

1 Use basic R commands to analyze biological data.

2 Apply statistics to solve problems in biological data.

3 Analyze gene expression results to find meaningful insights.

4 Evaluate RNA-Seq data to understand biological meaning.

Pre-requisite of course:NA

Teaching and Examination Scheme

Theory Hours	Tutorial Hours	Practical Hours	ESE	IA	CSE	Viva	Term Work
0	2	0	0	0	0	50	50

Contents : Unit	Topics	Contact Hours
Total Hours		

Suggested List of Experiments:

Contents : Unit	Topics	Contact Hours
1	**MODULE 1** Introduction, Setting up R, Basic functions in R, Basic data structures in R, Data structures and File I/O	4
2	**MODULE 2** Basic statistical tests with R, Correlation analysis, Analysis of variance (ANOVA), Basic data visualization techniques, Data visualization	4
3	**MODULE 3** Visualization with vioplot and ggplot2, R packages for plotting and data organization, Data transformation in R, Bioconductor packages, Flow cytometry data analysis in R/Bioconductor	4

DR. SEJAL SAHILKRISHNA SHAH
Digitally signed by (Name of HOD)

DR. RAJENDRASINH BAHADURSINH JADEJA
Digitally signed by (Name of Dean/ Principal)

Suggested List of Experiments:

Contents : Unit	Topics	Contact Hours
4	**MODULE 4** Gene expression analysis and co-expression network, WGCNA package and Data Download, WGCNA hands-on: Data preprocessing, WGCNA hands-on: Soft-threshold, WGCNA: Module gene expression	5
5	**MODULE 5** Introduction to ChIP-seq, ChIP-seq data analysis, ChIP-seq data analysis: Peak calling, Peak calling and Visualization, ChIP-seq data analysis: bigWig/bw files	5
6	**MODULE 6** Regression models on Biological data, Predictive models with linear regression, Multicollinearity, Lasso regression, Non-linear regression	4
7	**MODULE 7** Dimensionality reduction, Principal Component Analysis (PCA), PCA analysis hands-on, PCA analysis using 'PCAtools', UMAP analysis	4
Total Hours		**30**

Textbook :

1 Bioinformatics Data Skills: Reproducible and Robust Research with Open Source Tools, Vince Buffalo, O'Reilly Media, 2015

2 The Art of R Programming: A Tour of Statistical Software Design, Norman Matloff, Manning Publications, 2011

3 RNA-seq Data Analysis A Practical Approach, Eija Korpelainen, Chapman & Hall, 2015

References:

1 Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks, Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks, Trapnell, C., Roberts, A., Goff, L. et al, Nat Protoc , 2012

2 RNA-Seq Data Analysis: A Practical Guide for Model and Non-Model Organisms, RNA-Seq Data Analysis: A Practical Guide for Model and Non-Model Organisms, Pola-Sánchez, E., Hernández-Martínez, K. M., Pérez-Estrada, R., Sélem-Mójica, N., Simpson, J., Abraham-Juárez, M. J., ... & Villalobos-Escobedo, J. M.,, Current Protocols, 2024

3 An R package for weighted correlation network analysis, An R package for weighted correlation network analysis, Langfelder, P., & Horvath, S., BMC Bioinformatics, 2008

4 Cytoscape: a software environment for integrated models of biomolecular interaction networks, Cytoscape: a software environment for integrated models of biomolecular interaction networks, Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., ... & Ideker, T., Genome Research, 2003

5 edgeR: a Bioconductor package for differential expression analysis of digital gene expression data, edgeR: a Bioconductor package for differential expression analysis of digital gene expression data, Robinson, M. D., McCarthy, D. J., & Smyth, G. K., Bioinformatics, 2010

DR. SEJAL SAHILKRISHNA SHAH
Digitally signed by (Name of HOD)

DR. RAJENDRASINH BAHADURSINH JADEJA
Digitally signed by (Name of Dean/ Principal)

References:

6 Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2, Love, M. I., Huber, W., & Anders, S.v, Genome Biology, 2014

Suggested Theory Distribution:

The suggested theory distribution as per Bloom's taxonomy is as follows. This distribution serves as guidelines for teachers and students to achieve effective teaching-learning process

Distribution of Theory for course delivery					
Remember / Knowledge	**Understand**	**Apply**	**Analyze**	**Evaluate**	**Higher order Thinking / Creative**
0.00	0.00	35.00	35.00	30.00	0.00

Instructional Method:

1 The course delivery method will depend upon the requirement of content and need of students. The teacher in addition to conventional teaching method by blackboard, may also use any of tools such as demonstration, role play, Quiz, brainstorming, MOOCs etc.

2 The internal evaluation will be done on the basis of continuous evaluation of students in the laboratory and class-room.

3 Practical examination will be conducted at the end of semester for evaluation of performance of students in laboratory.

Supplementary Resources:

1 https://bioconductor.org/packages/release/bioc/html/DESeq2.html

2 https://bioconductor.org/packages/release/bioc/html/edgeR.html

3 https://onlinecourses.nptel.ac.in/noc25_bt43/preview

DR. SEJAL SAHILKRISHNA SHAH
Digitally signed by (Name of HOD)

DR. RAJENDRASINH BAHADURSINH JADEJA
Digitally signed by (Name of Dean/ Principal)