Bioconductor is an open software development for computational biology and bioinformatics. This is an open-source, open-development software project for the computational biology community. Github : https://github.com/Bioconductor.

It's open-source because anybody can read and modify the underlying code, and open-development because anybody can contribute and participate in the development of the code. It speaks the R language which is a widely used language for data science since it is a flexible language, especially for data analysis.

Most Bioconductor components are distributed as R packages. The functional scope of Bioconductor packages includes the analysis of DNA microarray, sequence, flow, SNP, and other genetic or genomic data. An R package is a structured collection of code (R, C, or other), documentation, and/or data for performing particular types of analysis, e.g., affy, cluster, graph packages. It provides executions of specific statistical and graphical methods.

The Bioconductor project provides access to powerful statistical and graphical methods for the analysis of genomic data. Analysis packages address workflows for analysis of oligonucleotide arrays, sequence analysis, flow cytometry. and other high-throughput genomic data. The R package system itself provides implementations for a broad range of state-of-the-art statistical and graphical techniques, including linear and non-linear modeling, cluster analysis, prediction, resampling, survival analysis, and time-series analysis.

A basic task in the analysis of count data from RNA-seq is the detection of differentially expressed genes. The count data are presented as a table which reports, for each sample, the number of sequence fragments that have been assigned to each gene. Analogous data also arise for other assay types, including comparative ChIP-Seq, HiC, shRNA screening, and mass spectrometry. An important analysis question is the quantification and statistical inference of systematic changes between conditions, as compared to within-condition variability. The package DESeq2 provides methods to test for differential expression by use of negative binomial generalized linear models; the estimates of dispersion and logarithmic fold changes incorporate data-driven prior distributions. This vignette explains the use of the package and demonstrates typical workflows. An RNA-seq workflow on the Bioconductor website covers similar material

to this vignette but at a slower pace, including the generation of count matrices from FASTQ files. DESeq2 package version: 1.42.0