

# The Biostar Handbook: A Beginner's Guide to Bioinformatics[7]

#course/biostar

## heatmap

`heatmap` is used to visualize differential expression genes for RNA-Seq and genomic features (promoters, CpG islands, enhancers, and gene bodies) for ChIP-seq signal.

### Posts:

- Two heatmap functions: `heatmap` in `stats` package  
`heatmap.2` in `gplots` package

RPosts - Understanding heatmaps, a tale of two heatmap functions

- You probably don't understand heatmaps |
- Clustering Then Scaling Or Scaling Then Clustering
- Scale Data Before Drawing Heatmap Or Using Heatmap(..., Scale="Column") In R?

### ?heatmap

scale character indicating if the values should be centered and scaled in either the row direction or the column direction, or none. **The default is "row"** if `symm` false, and "none" otherwise.

### symm

logical indicating if x should be treated symmetrically; can only be true when x is a square matrix.

### ?heatmap.2

scale character indicating if the values should be centered and scaled in either the row direction or the column direction, or none. **The default is "none"**.

**Scale parameters are different in them.**

"The defaults of almost every heat map function in R does the hierarchical clustering first, then scales the rows then displays the image"

**Scaling before clustering do make sense in the case of this post.**

This confirms that **heatmap** does clustering first, and then scale the row for representing the color

**Scale before clustering or use a different distance measure (correlation) can get the same clustering results. Scale parameter in heatmap function can affect the color of the output image.**

- Distance measure and linkage method can drastically affect clustering, so choose the right one.

[RPods - Heatmap demystified](#)

Genes are clustered on the basis of Pearson correlation. Samples are clustered on the basis of a Euclidian distance matrix with complete linkage.

The reason of using complete linkage is : it seems to outperform average linkage.

Euclidean distance, Pearson correlation and 'uncentered' correlation (angular separation) all seem to work reasonably well as distance measures. Euclidean distance may be more appropriate for log ratio data, whereas Pearson correlation seems to work better for absolute-valued (e.g., Affymetrix) data.

**While clustering rows and columns, use different distance measure(Euclidean distance, Pearson correlation or 'uncentered' correlation) and same linkage method(complete).**

[Visualization of ChIP-seq data using Heatmaps \(Updated: 06/10/16\)](#)

Paper:



heatmap-p1.pdf  
Apr 5, 2019

**Mapping quantitative data to color**



heatmap-p2.pdf  
Apr 5, 2019

## Heat maps



heatmap-p3.pdf  
Apr 8, 2019

## Low-coverage single-cell mRNA sequencing reveals cellular heterogeneity and activated signaling pathways in developing cerebral cortex



heatmap-p4.pdf  
Apr 8, 2019

## How does gene expression clustering work?

### Tools:

[heatmap.2](#)

[pheatmap](#)

[heatmap](#)

[GitHub - jokergoo/ComplexHeatmap: Make Complex Heatmaps](#)

[Bioconductor - EnrichedHeatmap](#)

[Bioconductor - ComplexHeatmap](#)

### GUI tools:

[EaSeq - Interactive ChIP-seq analysis software](#)