Heatmap

```r
library(DESeq2)
library(pheatmap)

## Load count matrix and metadata

counts <- read.csv("C:/Users/valfr/OneDrive/Desktop/counts_matrix.csv", row.names = 1)
coldata <- read.csv("C:/Users/valfr/OneDrive/Desktop/metadata.csv", row.names = 1)

## Create DESeq2 object and run DESeq

ds <- DESeqDataSetFromMatrix(
  countData = counts,
  colData   = coldata,
  design    = ~ condition
)

ds <- DESeq(ds)

## Variance-stabilising transform

# turns raw counts into log-like values with stabilised variance,
# making samples easier for clustering / heatmaps.
vsd  <- vst(ds, blind = FALSE)

# Extract the expression matrix
expr <- assay(vsd)

## Order samples to match the paper

# Paper order: SY14_3, SY14_1, SY14_2, BY4742_2, BY4742_1, BY4742_3

sample_order <- c("SY14_R3","SY14_R1","SY14_R2",
                  "BY4742_R2","BY4742_R1","BY4742_R3")

expr     <- expr[, sample_order]
coldata <- coldata[sample_order, , drop = FALSE]

## Compute sample-sample Pearson correlations

# Correlation matrix between samples (columns of expr)
cor_mat <- cor(expr)
round(cor_mat, 3)
```

```
##           SY14_R3 SY14_R1 SY14_R2 BY4742_R2 BY4742_R1 BY4742_R3
## SY14_R3     1.000   0.997   0.997     0.986     0.985     0.986
```

```
## SY14_R1      0.997    1.000    0.998      0.985      0.985      0.986
## SY14_R2      0.997    0.998    1.000      0.986      0.986      0.986
## BY4742_R2    0.986    0.985    0.986      1.000      0.998      0.998
## BY4742_R1    0.985    0.985    0.986      0.998      1.000      0.998
## BY4742_R3    0.986    0.986    0.986      0.998      0.998      1.000
```
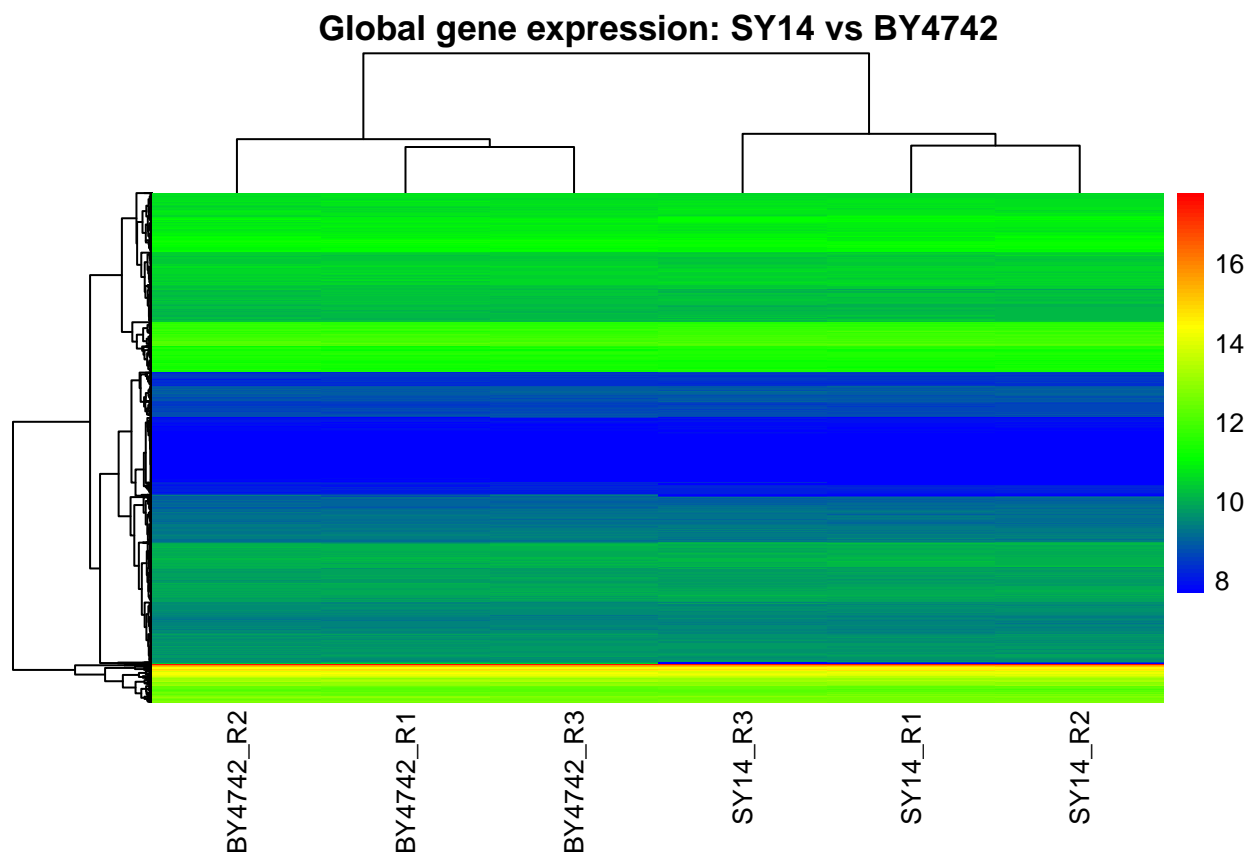
Sample–sample Pearson correlations were extremely high. Within-group correlations 0.997–0.998 for SY14 and 0.998 for BY4742. Between-group correlations were 0.985–0.986, very close to the values reported in the original study (>0.98 within groups, >0.97 between groups). This confirms that the global transcriptome profiles of SY14 and BY4742 are highly similar.

```r
## Plot heatmap

#use similar colour palette
my_colors <- colorRampPalette(c("blue","green","yellow","red"))(100)

val_range <- range(expr)
my_breaks <- seq(val_range[1], val_range[2], length.out = 101)

pheatmap(
  expr,
  color         = my_colors,
  breaks        = my_breaks,
  show_rownames = FALSE,    # hide thousands of gene IDs
  show_colnames = TRUE,
  cluster_rows  = TRUE,     # gene dendrogram (left)
  cluster_cols  = TRUE,     # sample dendrogram (top)
  border_color  = NA,
  fontsize_col  = 10,
  angle_col     = 90,
  main          = "Global gene expression: SY14 vs BY4742"
)
```

Global gene expression: SY14 vs BY4742

The heatmap displays the global transcriptome profiles of the six RNA-seq samples. Each row represents a gene, and each column represents a sample. The colours correspond to the variance stabilised expression values produced by DESeq2. The purpose of this heatmap is to assess how similar the overall gene-expression patterns are across samples and whether the two strains have globally distinct or similar transcriptomes.

Variance stabilising transformation is used because it corrects for differences in sequencing depth and stabilises the variance across the full range of gene expression, allowing fair comparison between samples.

Using VST values results in a wider expression range (around 8–16), which explains why this heatmap looks smoother and less colour variable than the original SY14 paper. The published figure used a different transformation which was not established and compresses values into a 0–4 range generating colour differences. The dendrogram shows that the three replicates of each strain cluster tightly together. The SY14 samples cluster very closely with the BY4742 samples, indicating that their global gene-expression profiles are highly similar.

In conclusion these results show that the fusion of all 16 yeast chromosomes into a single chromosome has little effect on global transcription as the transcriptomes of SY14 and BY4742 remain almost indistinguishable.