BB512/BB612 - Homework I

Mar 17, 2022

Use the `montpick` eset to perform the required analyses:

```
con <- url("http://bowtie-bio.sourceforge.net/recount/ExpressionSets/montpick_eset.RData")
load(file = con)
close(con)


pdata <- pData(montpick.eset)
edata <- exprs(montpick.eset)
fdata <- fData(montpick.eset)
```

Proprocessing, EDA and Clustering

1. [5 pt] Exclude probes with average expression count < 100

2. [5 pt] Perform log2 transformation

3. [10 pt] Perform quantile normalization, keeping the row and column names

4. [5 pt] Check the distributions via a boxplot

5. [15] Perform hierarchical clustering with average agglomeration (UPGMA) and plot the dendrogram (You may use any appropriate distance metric)

6. [15] Perform hierarchical clustering with average centroid (UPGMC) and plot the dendrogram (You may use any appropriate distance metric)

7. [10] Plot the dendrogram of UPGMA, coloring leaves by `population` (in `pdata`)

PCA

8. [15 pt] Perform PCA of samples, scaling the variables

9. [10 pt] Plot the scree plot, showing the percentage of variances explained by each principal component

10. [10 pt] Plot PC1 vs PC2 showing individuals colored by population