Figure 1: (A) A microarray may contain thousands of ‘spots’. Each spot contains many copies of the same DNA sequence that uniquely represents a gene from an organism. Spots are arranged in an orderly fashion into Pen-groups. (B) Schematic of the experimental protocol to study differential expression of genes. The organism is grown in two different conditions (a reference condition and a test condition). RNA is extracted from the experimental protocol to study differential expression of genes A to E could be arranged in any order and it is the task of the methods to arrange them meaningfully.

Figure 2: Zooming onto a spot on the microarray slide. The spot area and the background area are depicted by a blue circle and a white box, respectively. A pixel in the spot area is also shown. Any pixel within the blue circle will be treated as a signal from the spot. Pixels outside the blue circle but within the white box will be treated as a signal from the background. One can see that the images are not perfect, as it is often the case, which leads to many problems with spurious signals from dust particles, scratches, bright arrays, etc. This image was retrieved from SMD.

Figure 3: Gene expression data before and after the normalization procedure. Note that before normalization the image had many spots of different intensities, but after normalization only spots that are really different light up. This image was kindly provided by Lascombe, N.

Figure 4: A Schematic showing the principle behind agglomerative and divisive clustering. The colour code represents the log (expression ratio), where red represents up-regulation, green represents down-regulation, and black representing no change in expression. In aggregative clustering, genes that are similar to each other are grouped together, and an average expression profile is calculated for the group by using the average linkage algorithm. This step is performed iteratively until all genes are included into one cluster. In the case of divisive clustering, the whole set of genes is considered as a single cluster and is broken down iteratively into sub-clusters with similar expression profiles until each cluster contains only one gene. This information can be represented as a tree, where the terminal nodes represent genes and all branches represent different clusters. The distance from the branch point provides a measure of the distance between two objects. This image was adapted from Dopazo et al., (2001). Notice that the matrix at the top is the actual product of either aggregative or divisive clustering, and genes A to E are given in the final order for the simplicity of representation; initially rows corresponding to genes A to E could be arranged in any order and it is the task of the methods to arrange them meaningfully.