**Department of Computing and Information Systems**

**COMP 90016**

Workshop 10

In this workshop we will see an alternative method to CNV calling (see Workshop 6) based upon hierarchical clustering.

**Your tasks in this workshop**.

1. Work in an interactive shell in R. Use the binned read counts in assignment 3’s working directory on digitalis (mapped\_pairs.wig). Transform the bins into log2-ratios (log2(bin/mean)) – add a minimum offset to avoid 0 bins.
2. Create the distance matrix (dist(log-ratios))
3. Perform HC on matric (hclust(D))  
   Plot it (png(‘hclust.png’); plot(hc); dev.off())

How does this compare to the CBS exercise we did?  
What are the different clusters?   
When analysing CNVs, what cutoff height would be meaningful?  
- Extract one of the main cluster after cutting the tree and explore the data (summary(logfc[logfc$clusterno==X,”logfc”]))

1. Try a different linkage method and repeat.  
   How and why does the clustering change? What is a good method for the data? How could we transform this data into a 2-dimensional array (i.e., pos x bin size). How could we cluster intervals like this?

451

**Complete Linkage of (bin,logR) 2-dimensional clustering**

