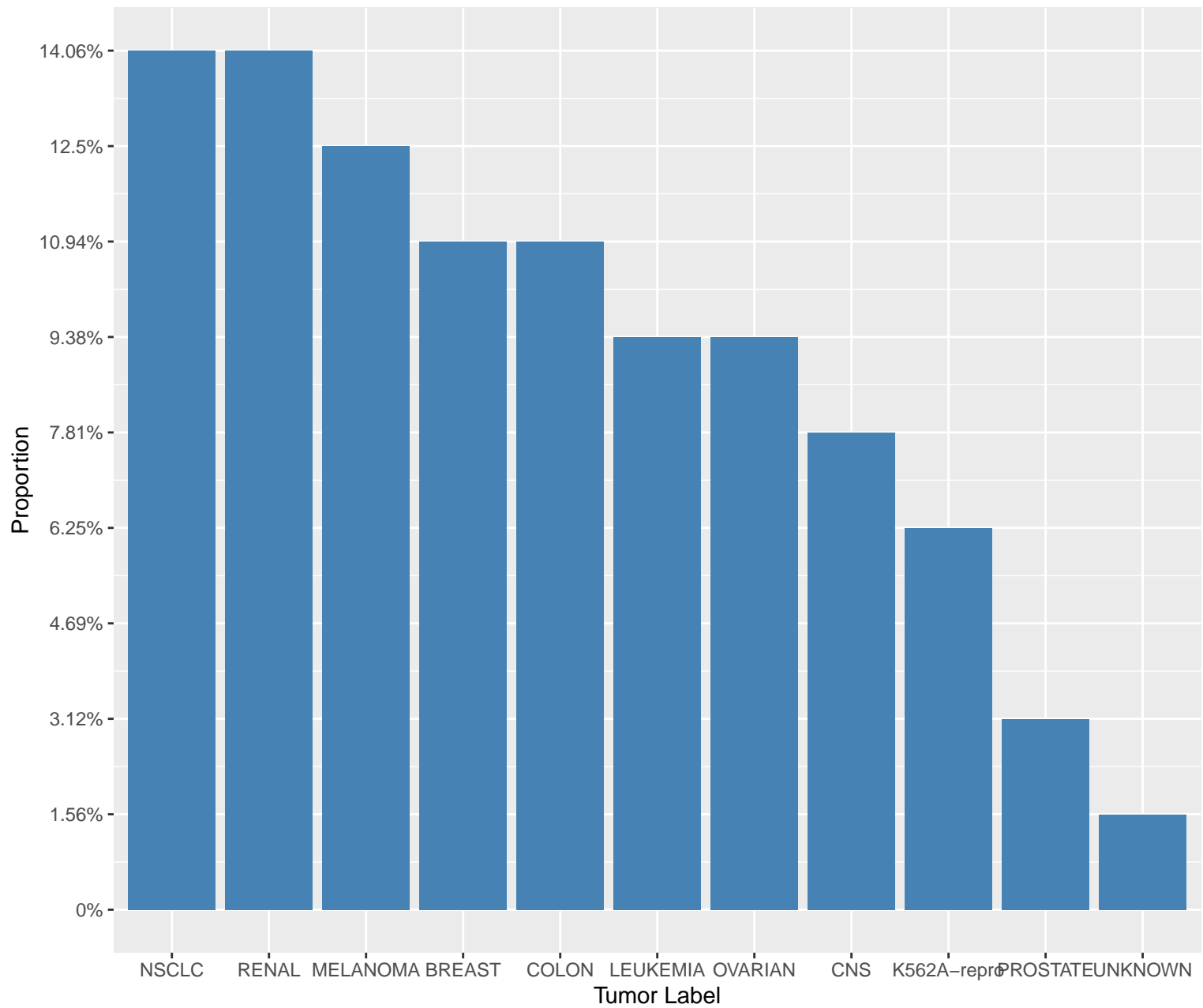
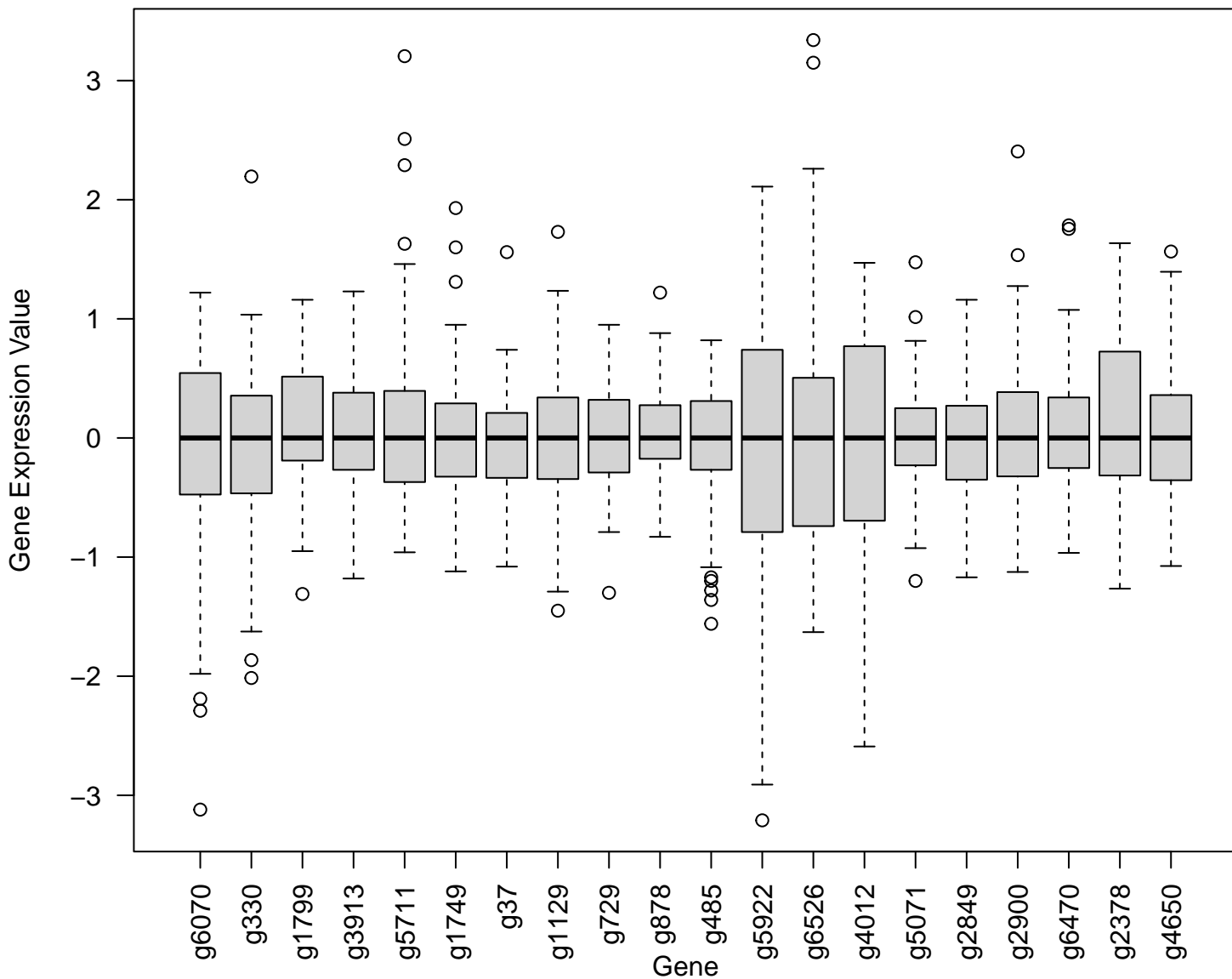


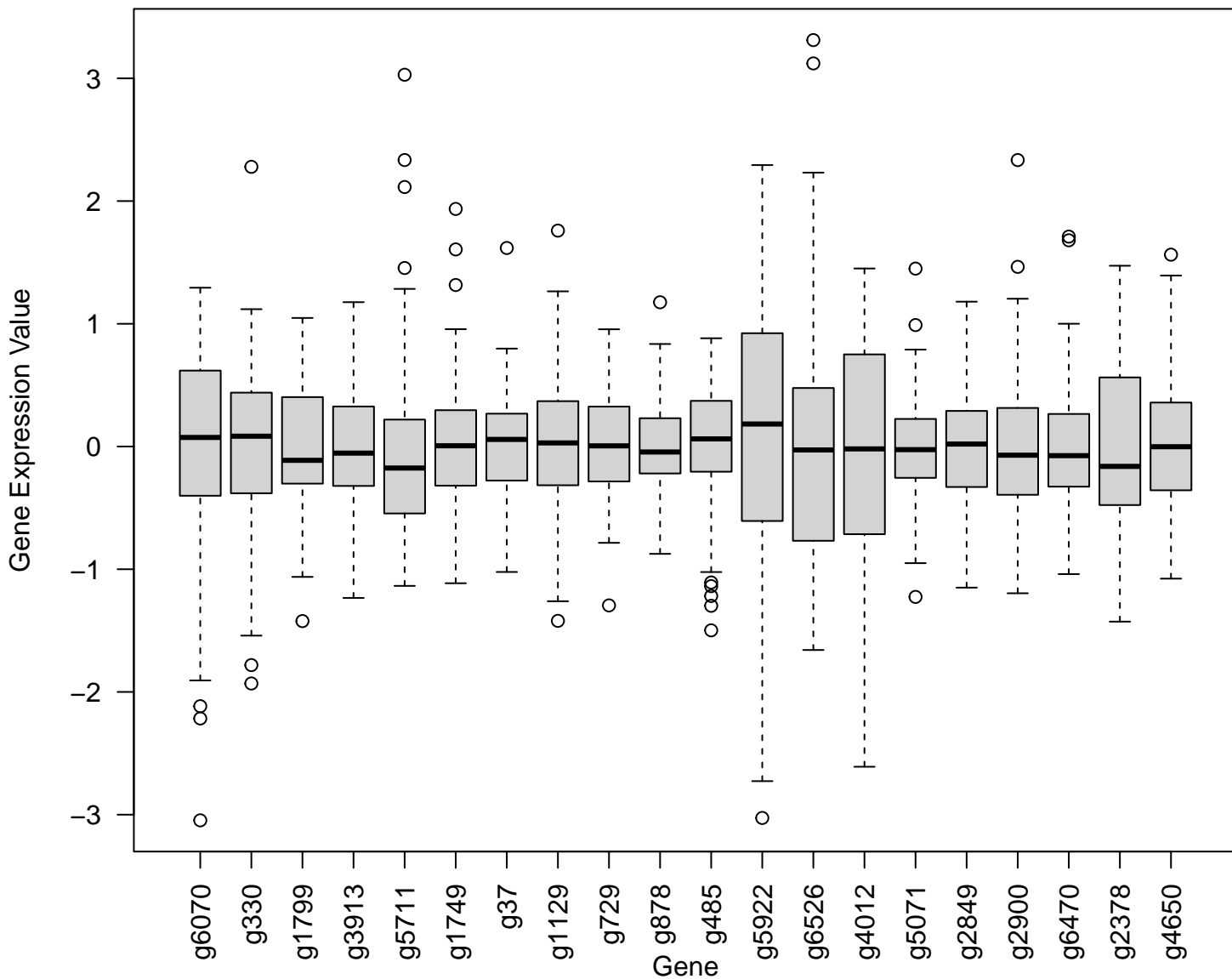
Tumor Label Distribution



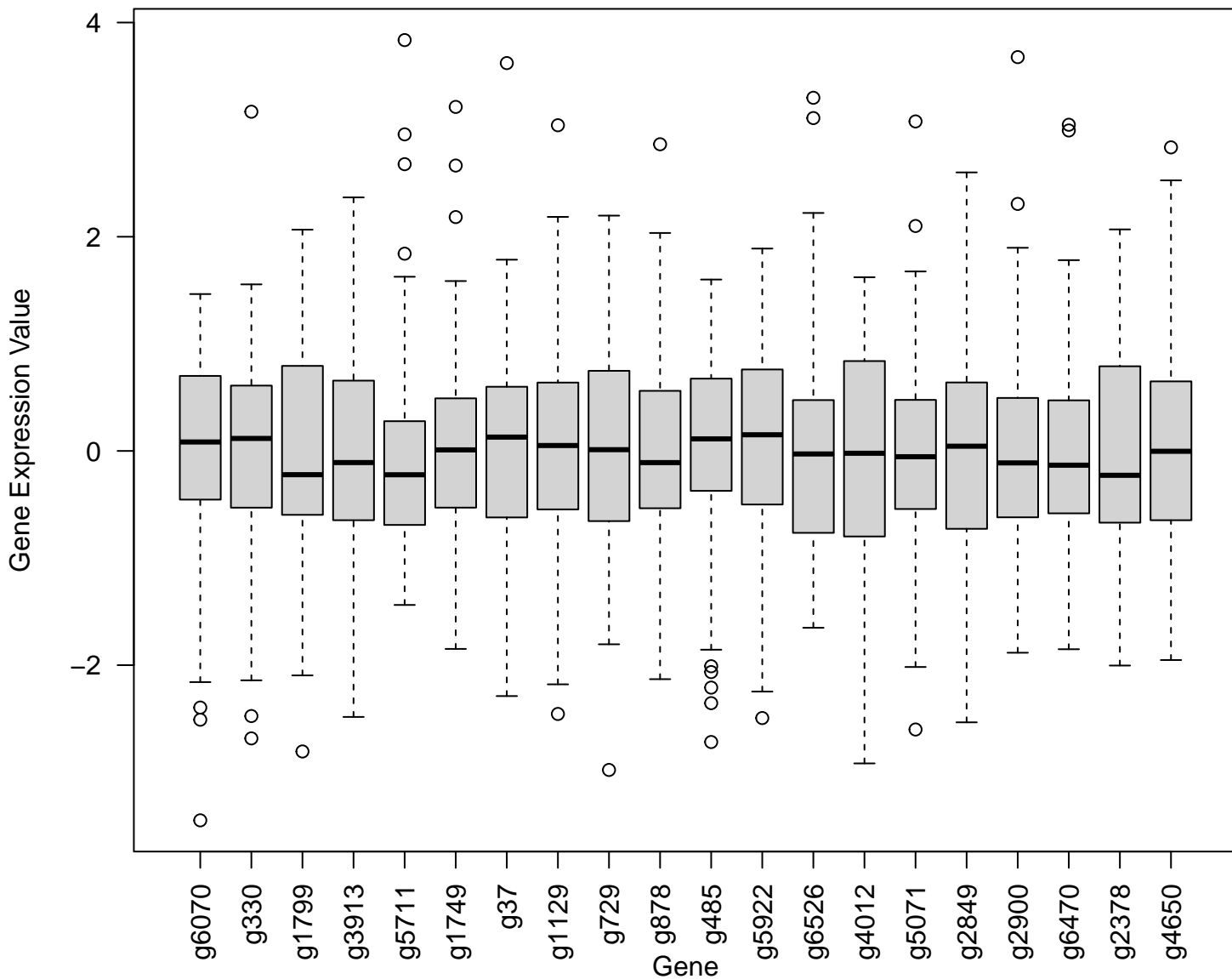
**Subset Gene Expression marginalised over patients**



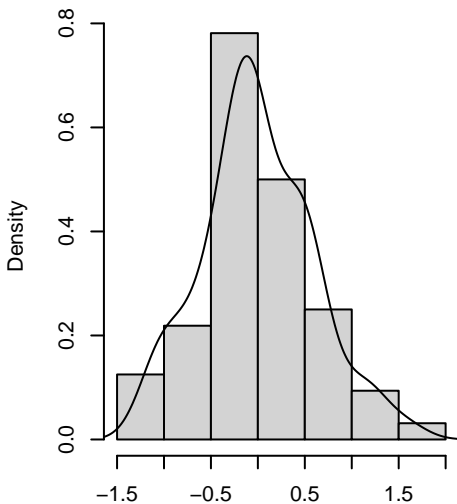
**Subset Gene Expression marginalised over patients (centered)**



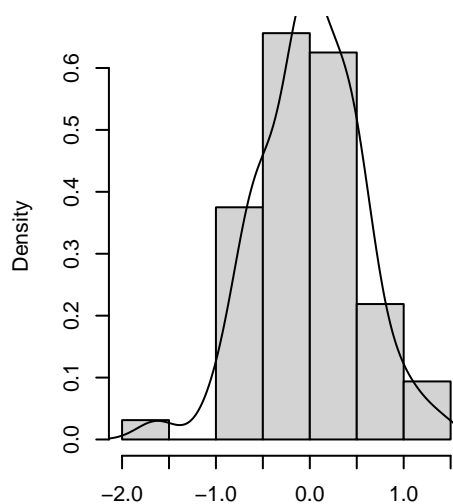
# Subset Gene Expression marginalised over patients (centered and scaled)



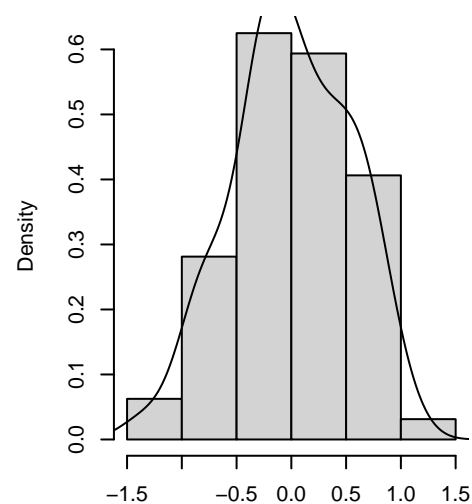
**g1533 Distribution**



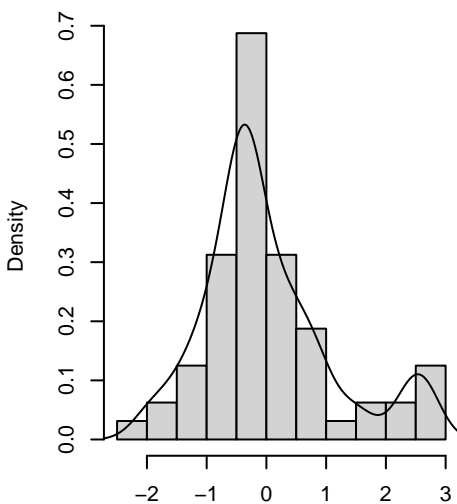
**g4567 Distribution**



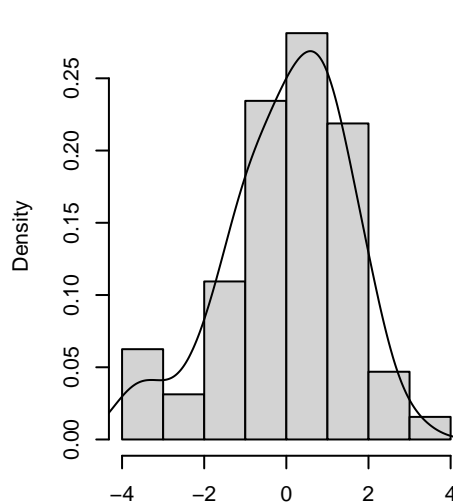
**g2347 Distribution**



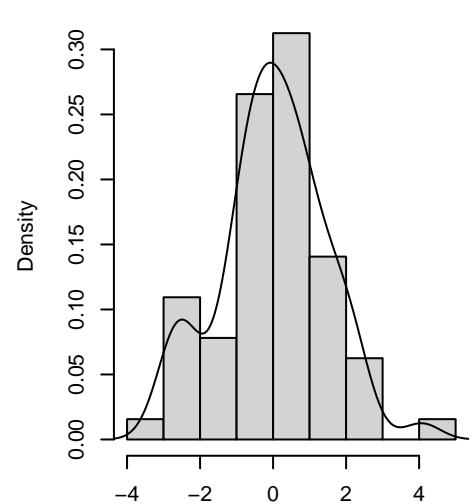
**g270 Distribution**

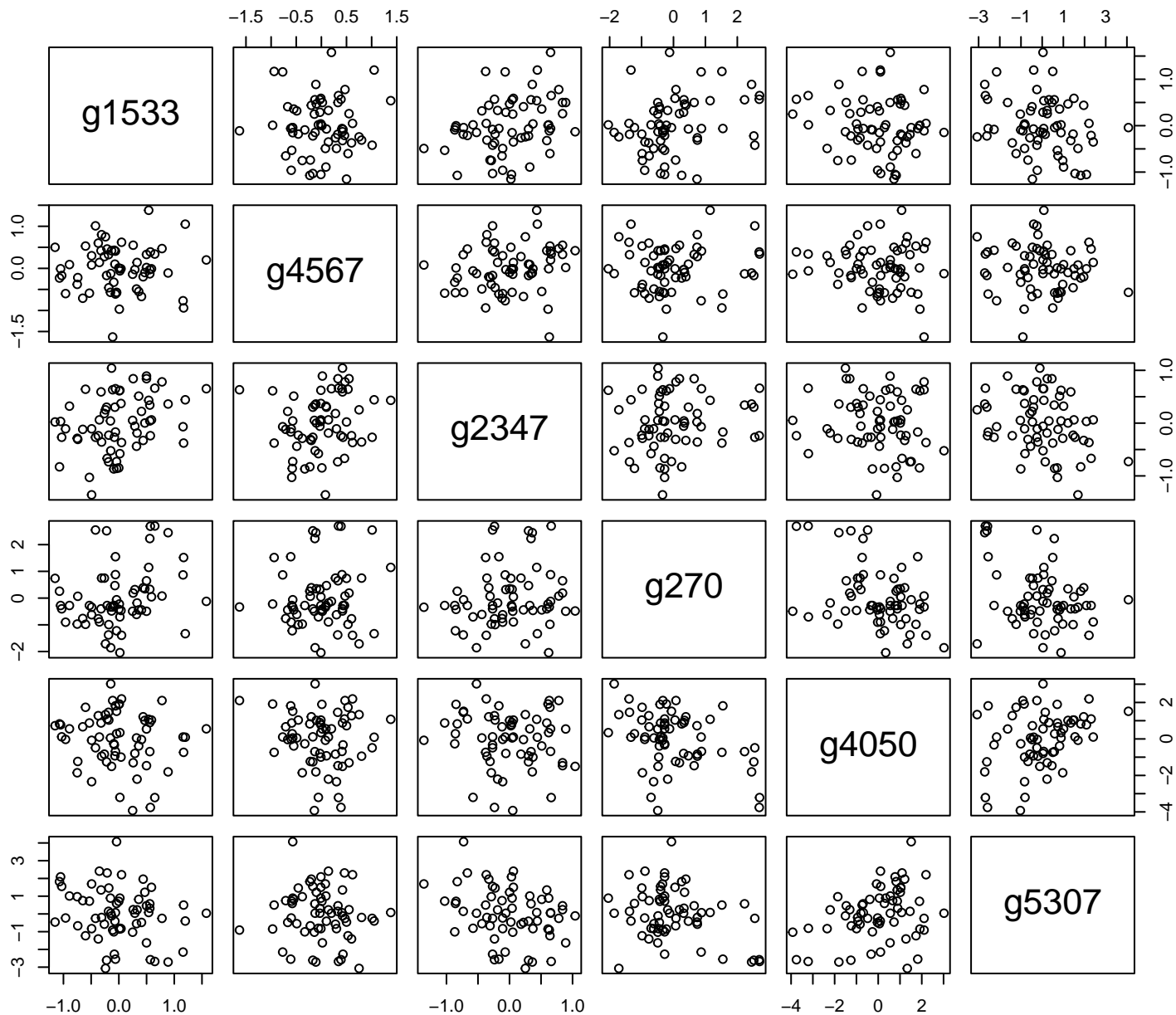


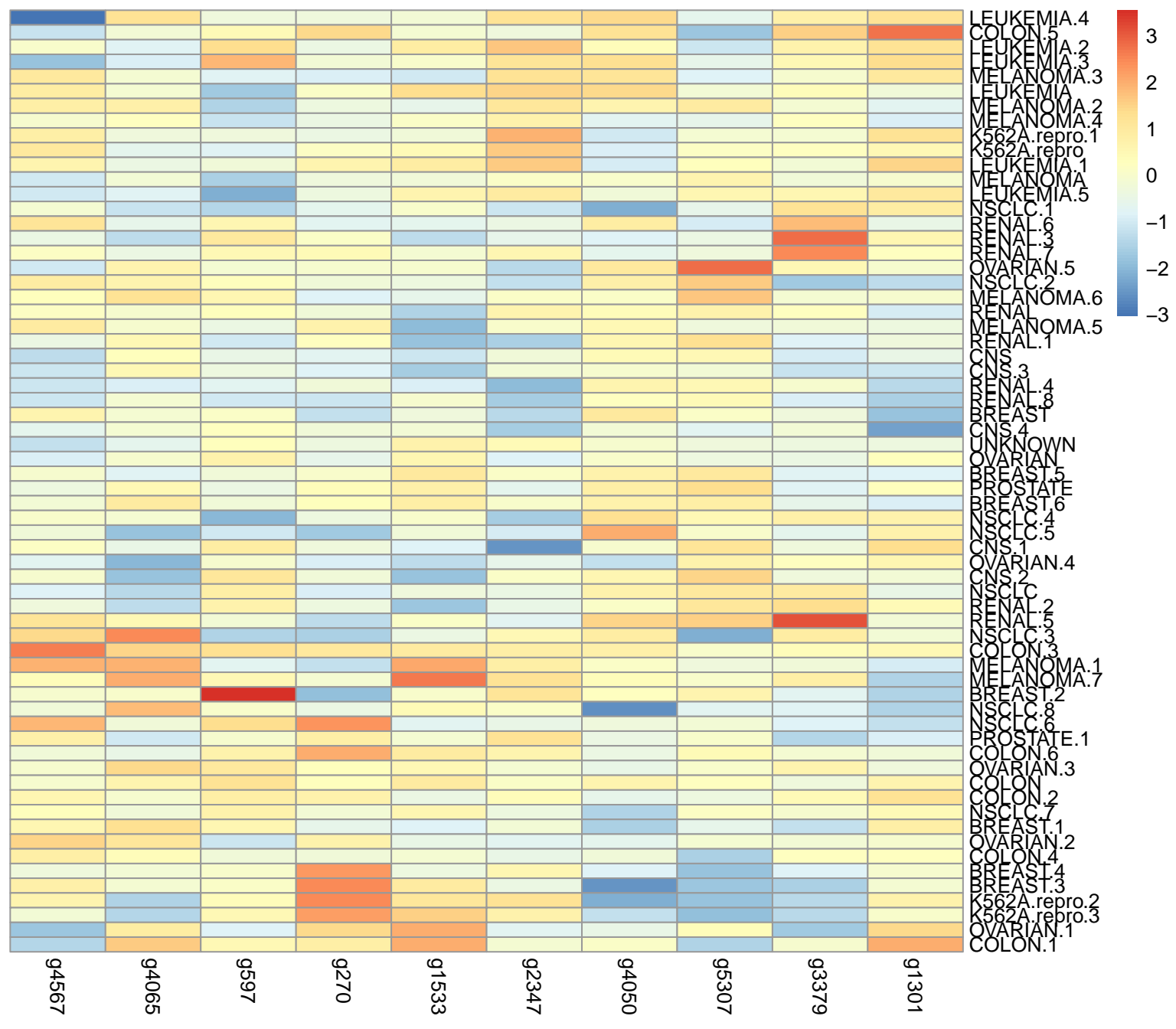
**g4050 Distribution**



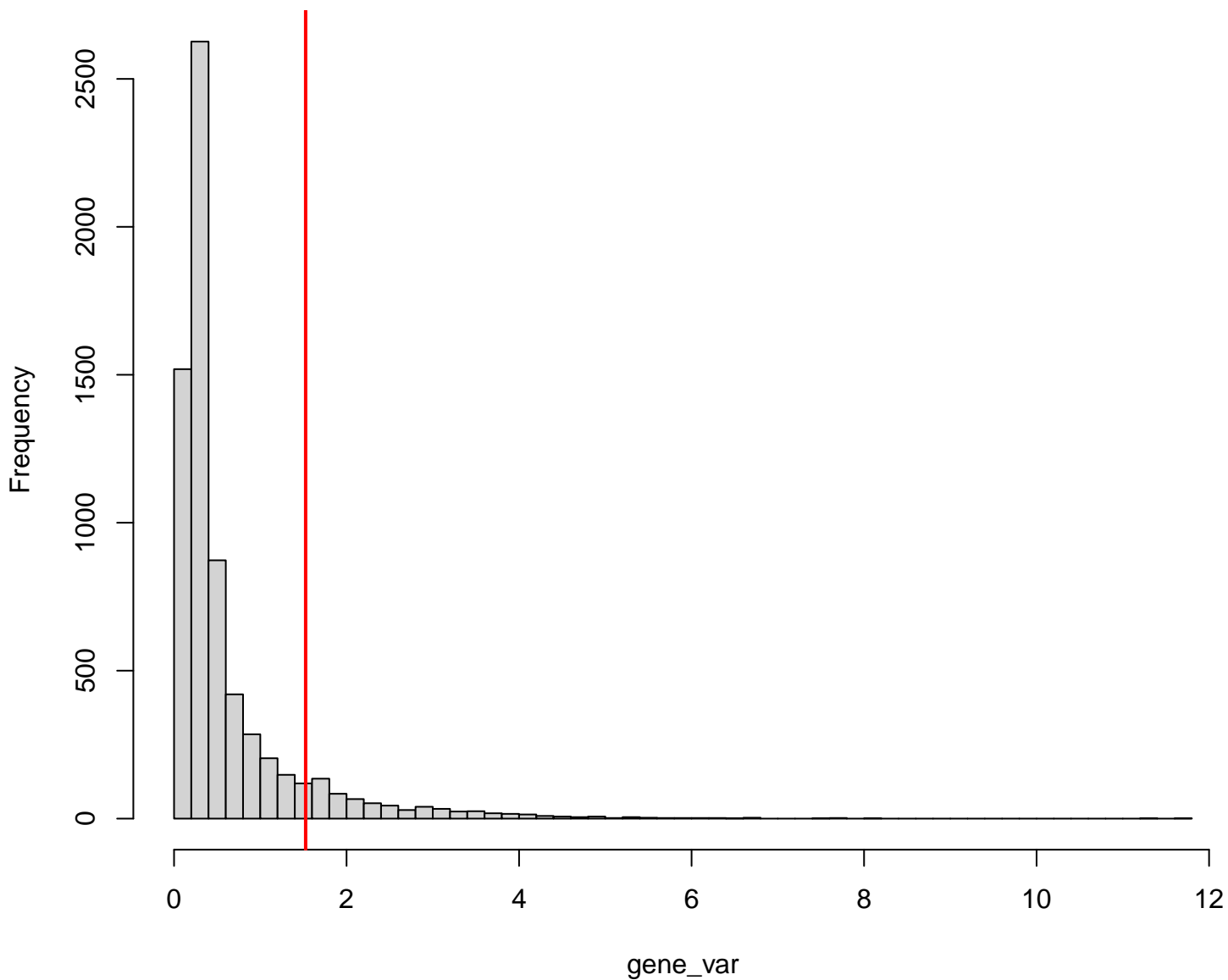
**g5307 Distribution**





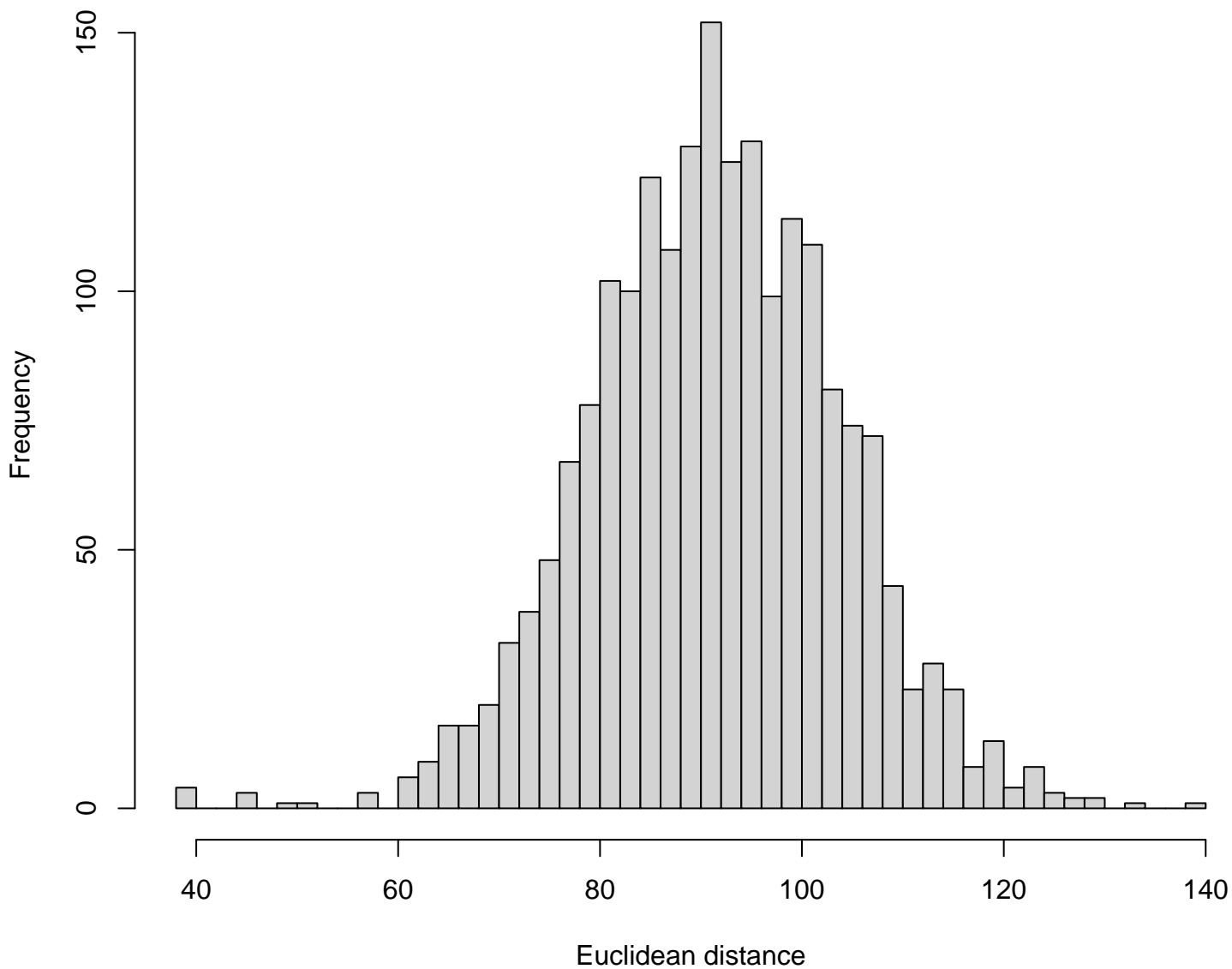


# Gene-wise variance distribution

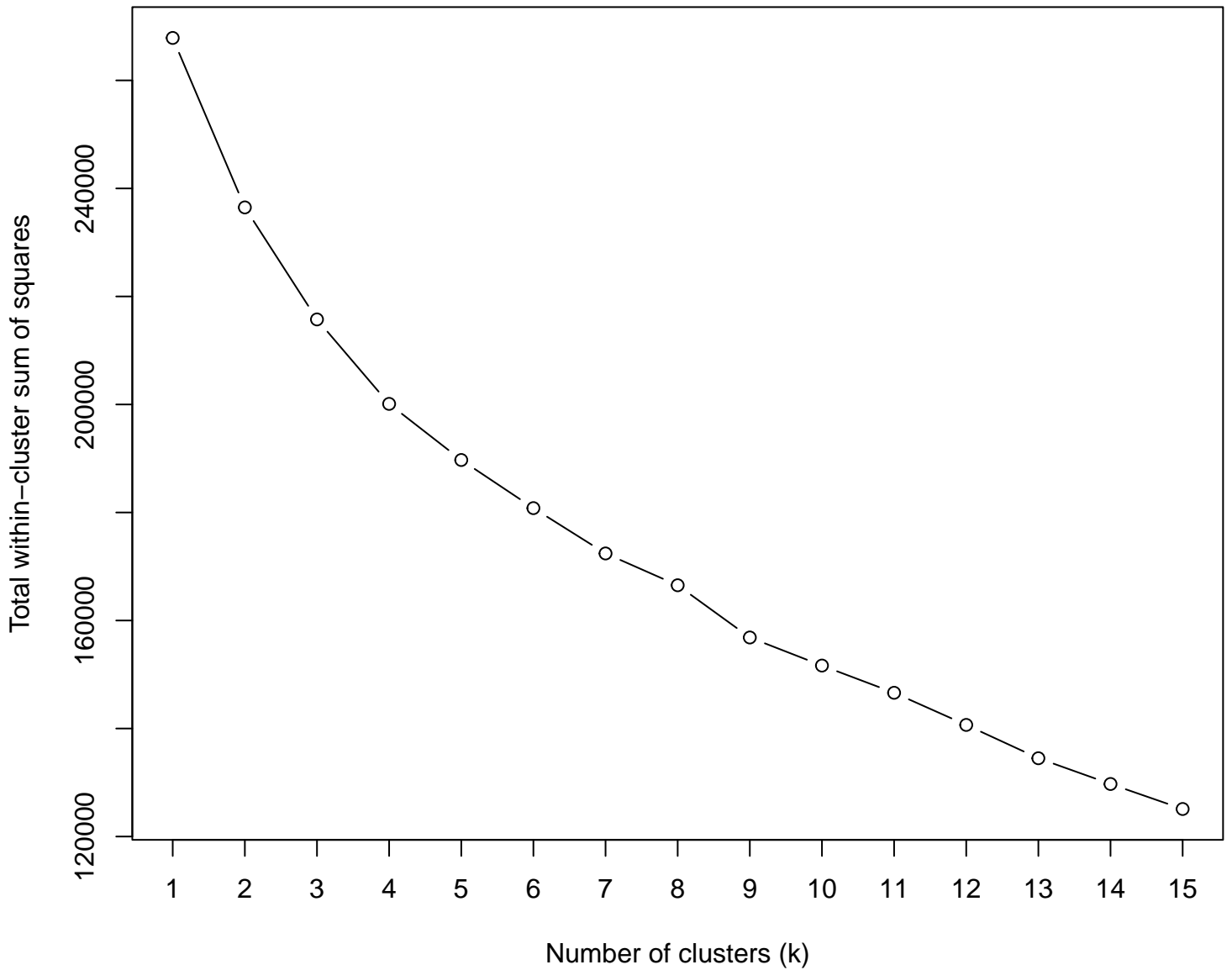




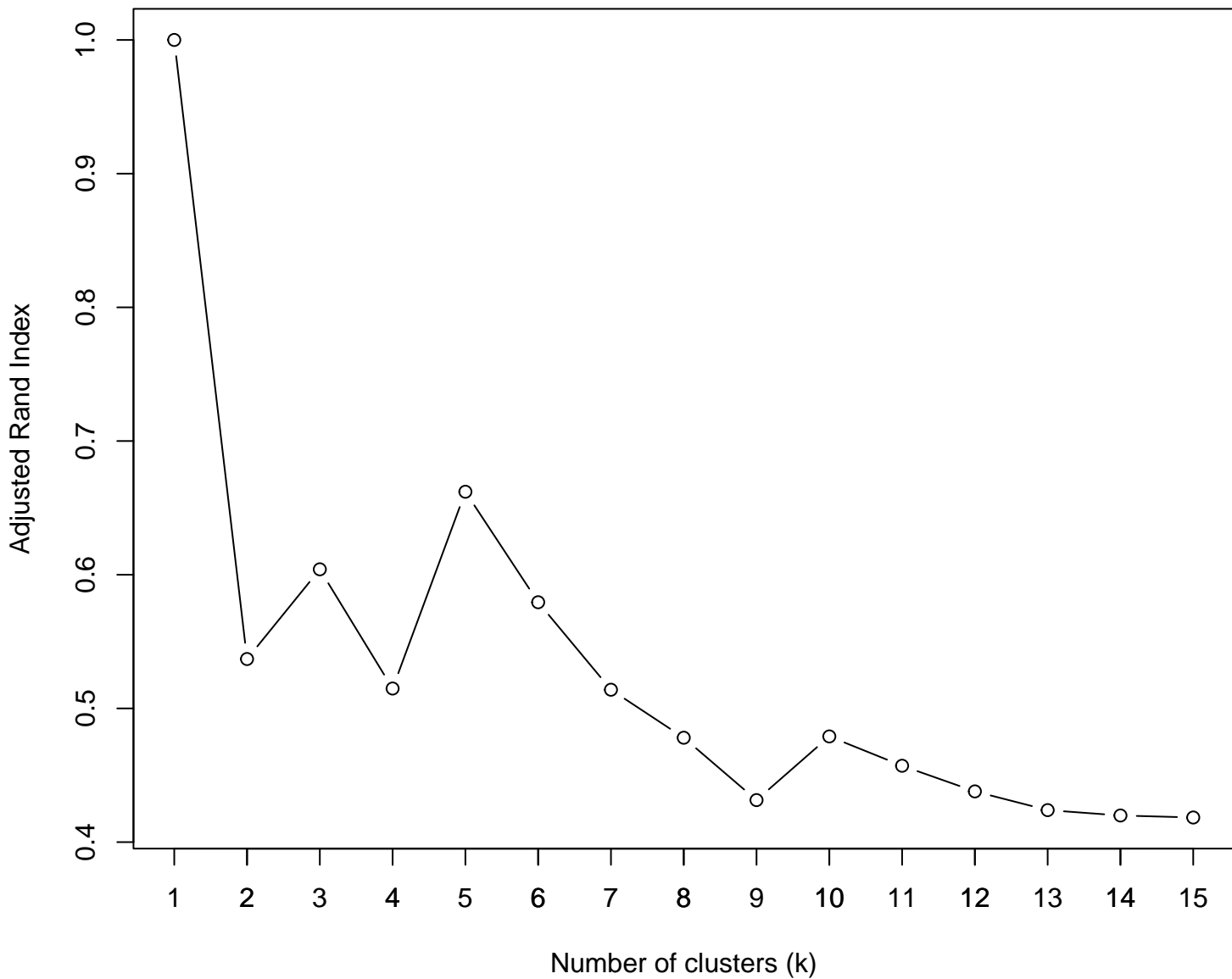
## Pairwise distances between samples



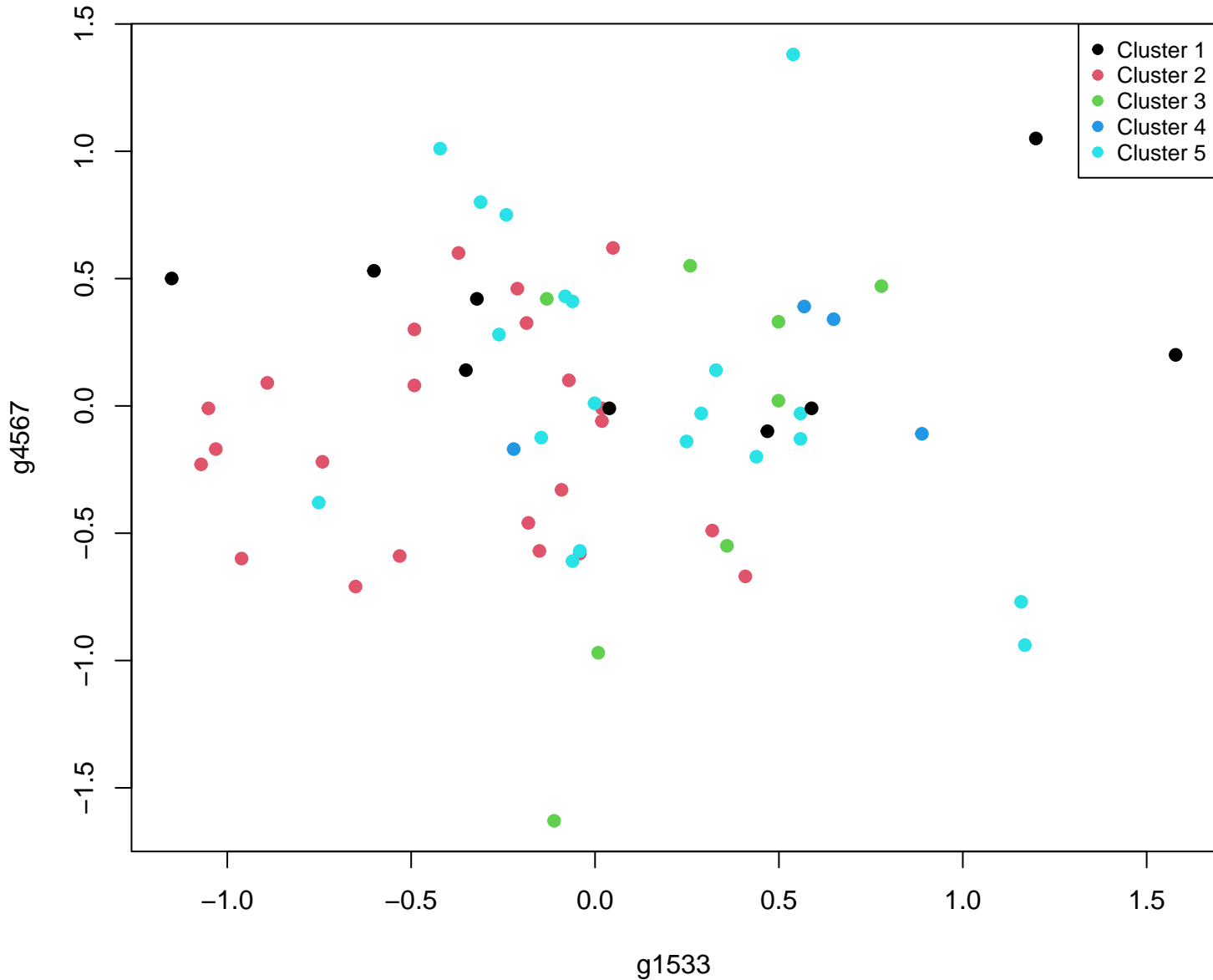
**Elbow plot for k-means clustering**



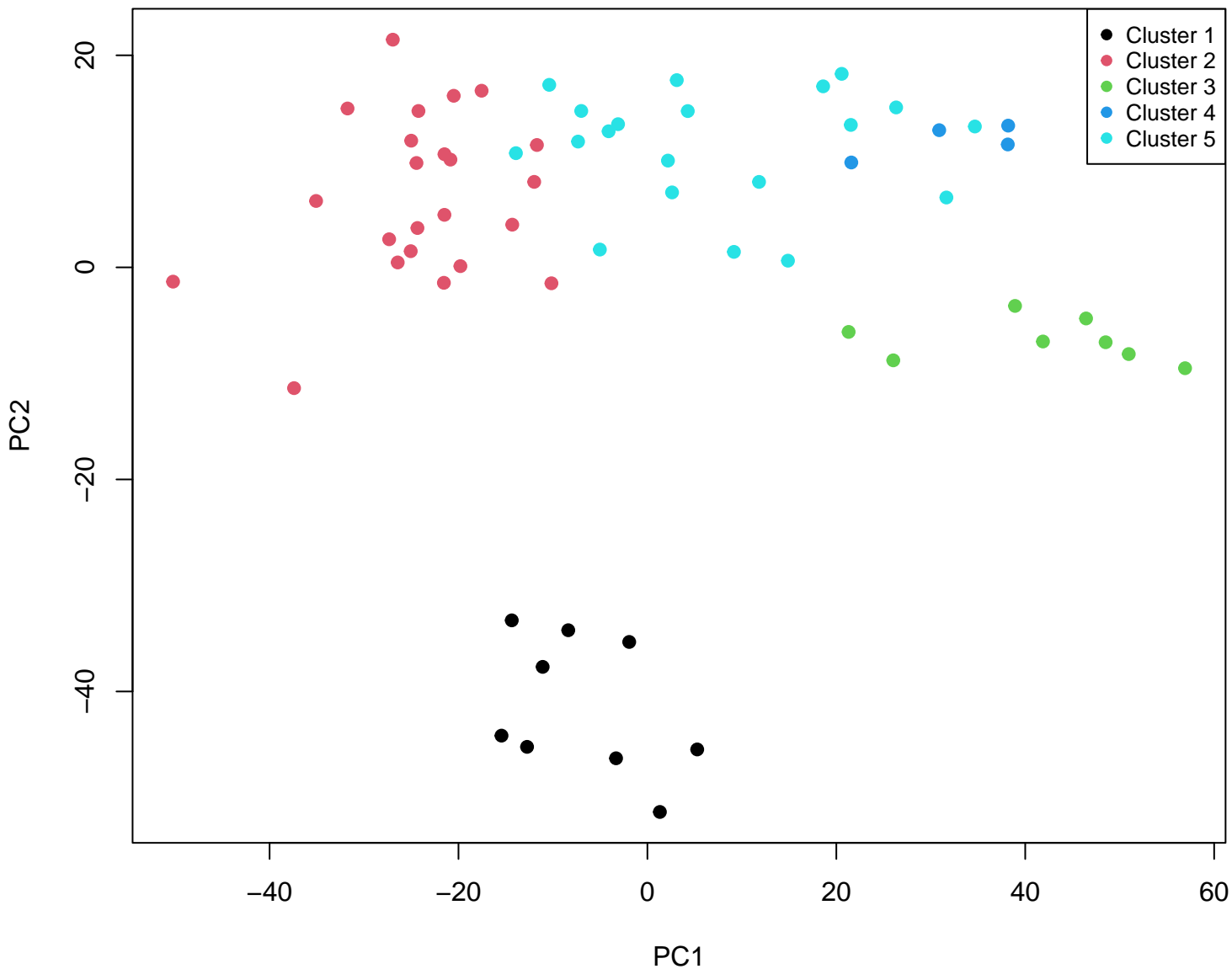
**ARI vs k clusters**



**k-means clusters with 5 clusters (using 2 randomly sampled genes)**

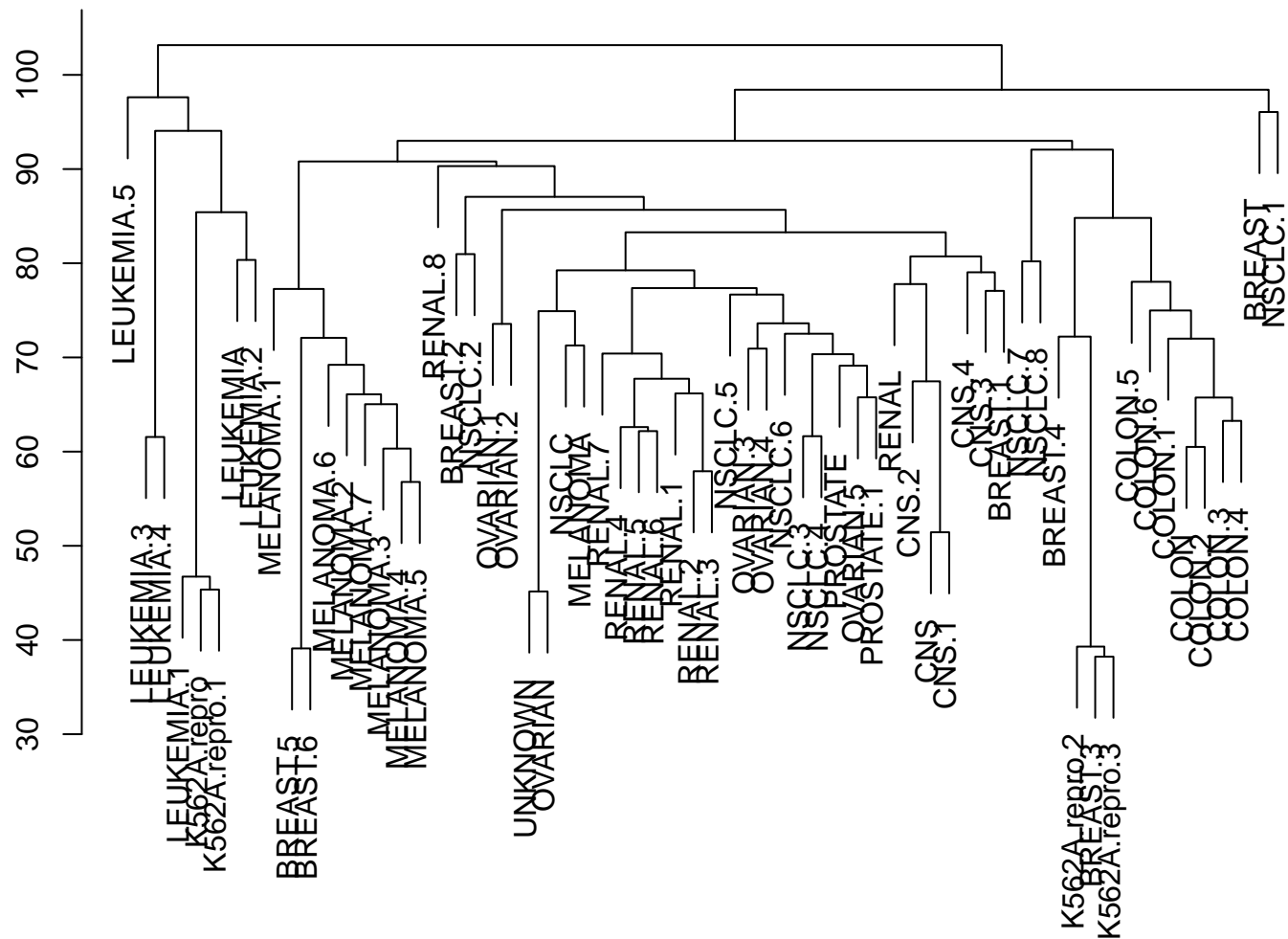


**K-means clusters (first 2 PCs, raw clustering)**



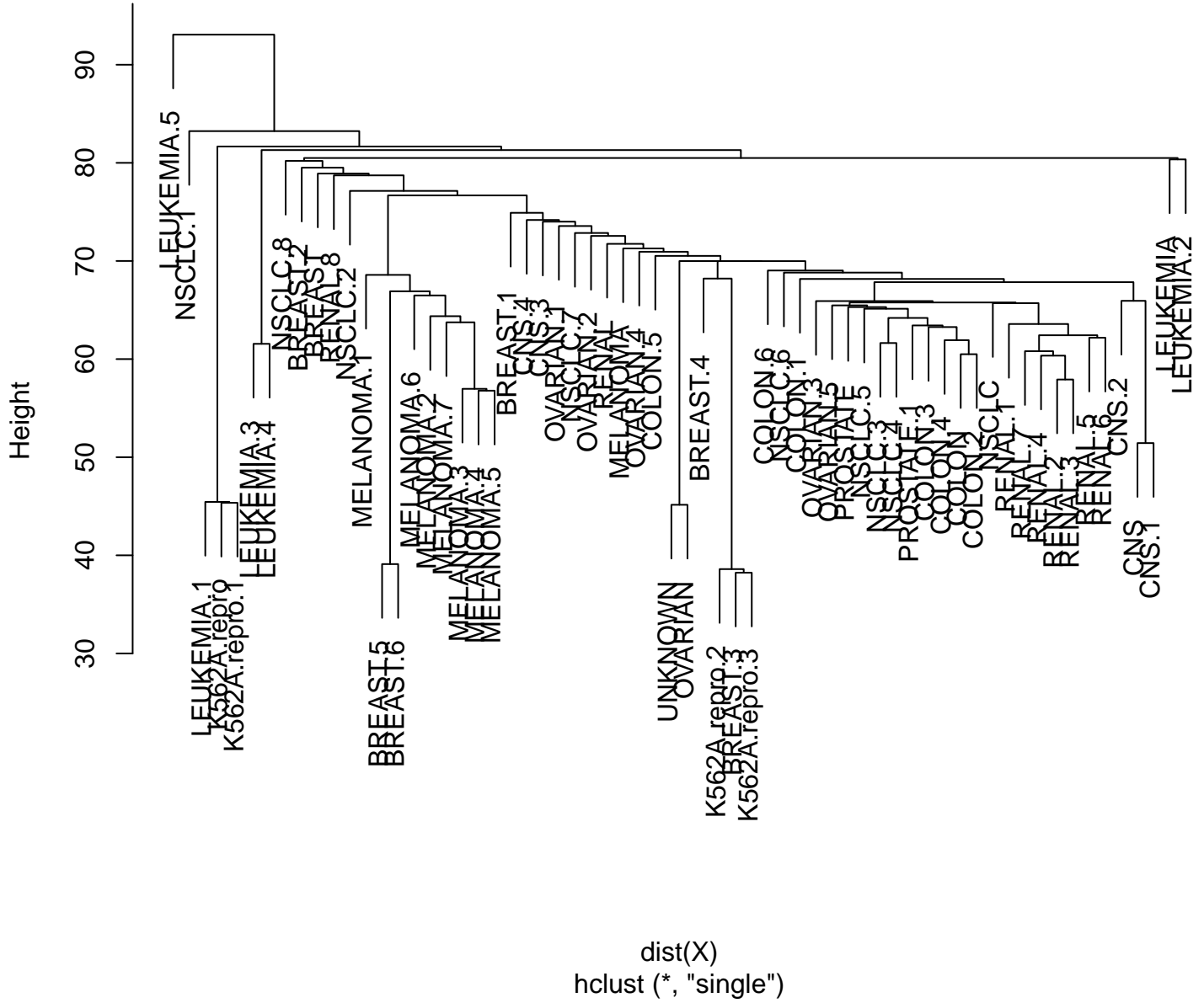
Height

Average

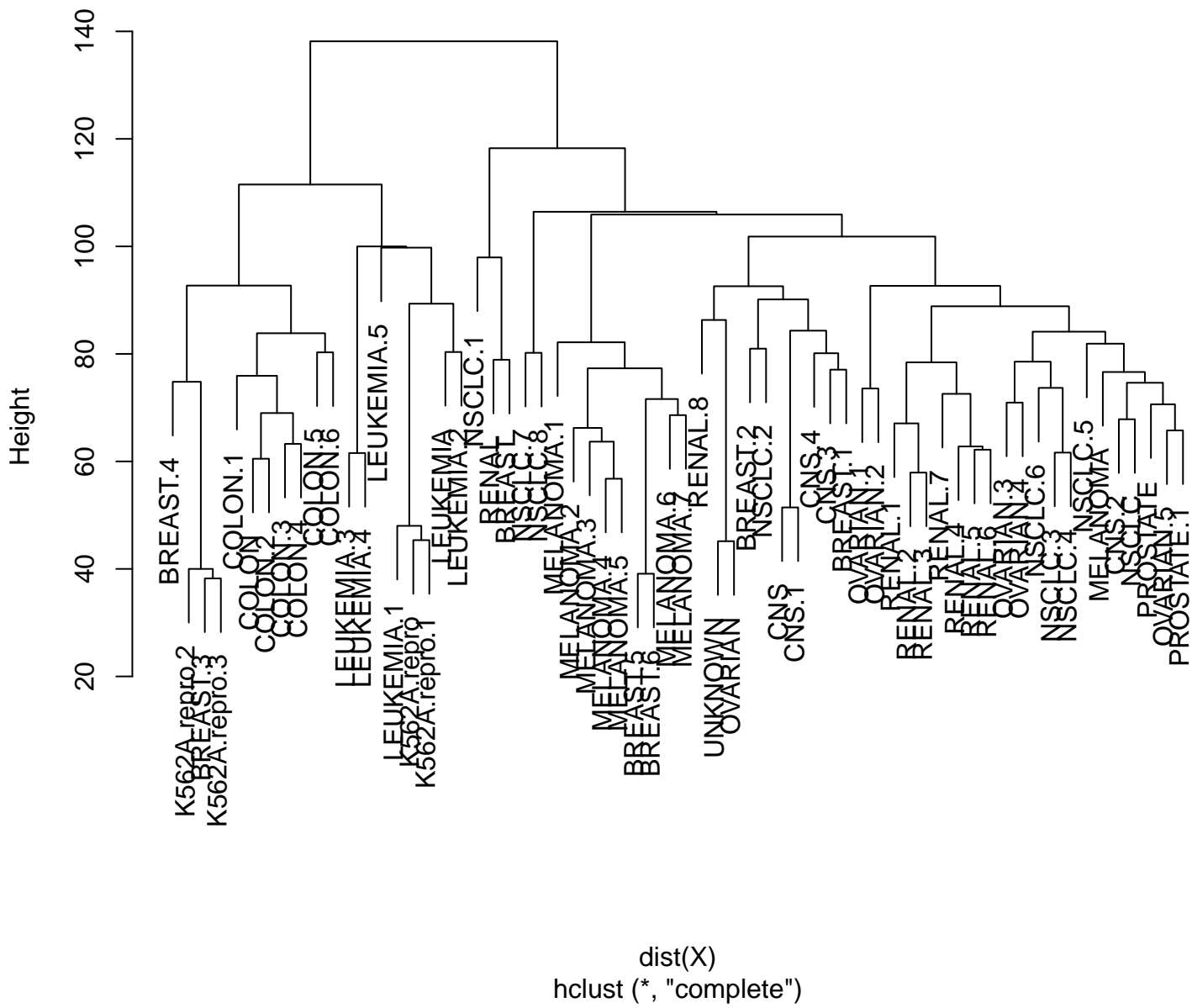


dist(X)  
hclust (\*, "average")

## Single

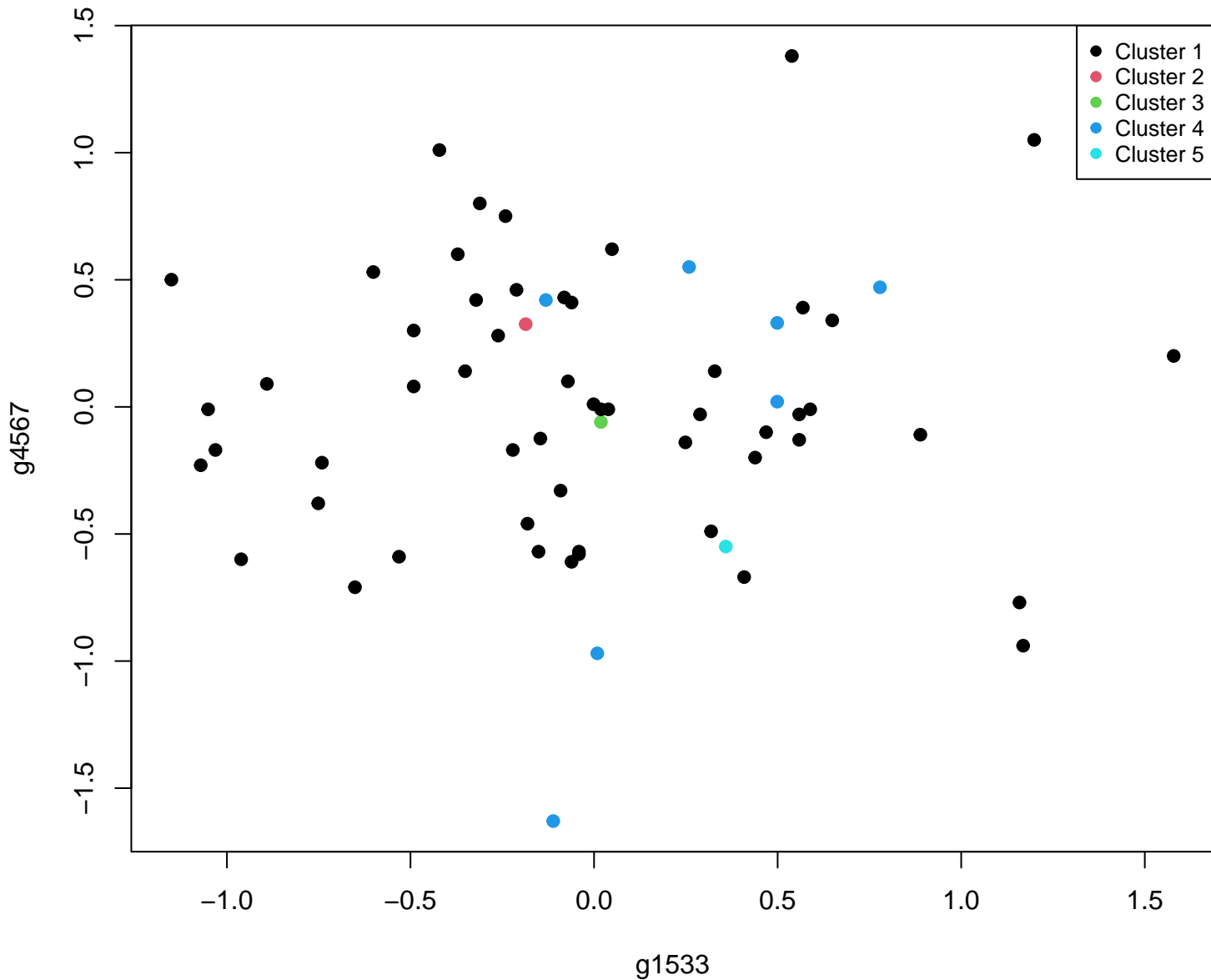


# Complete

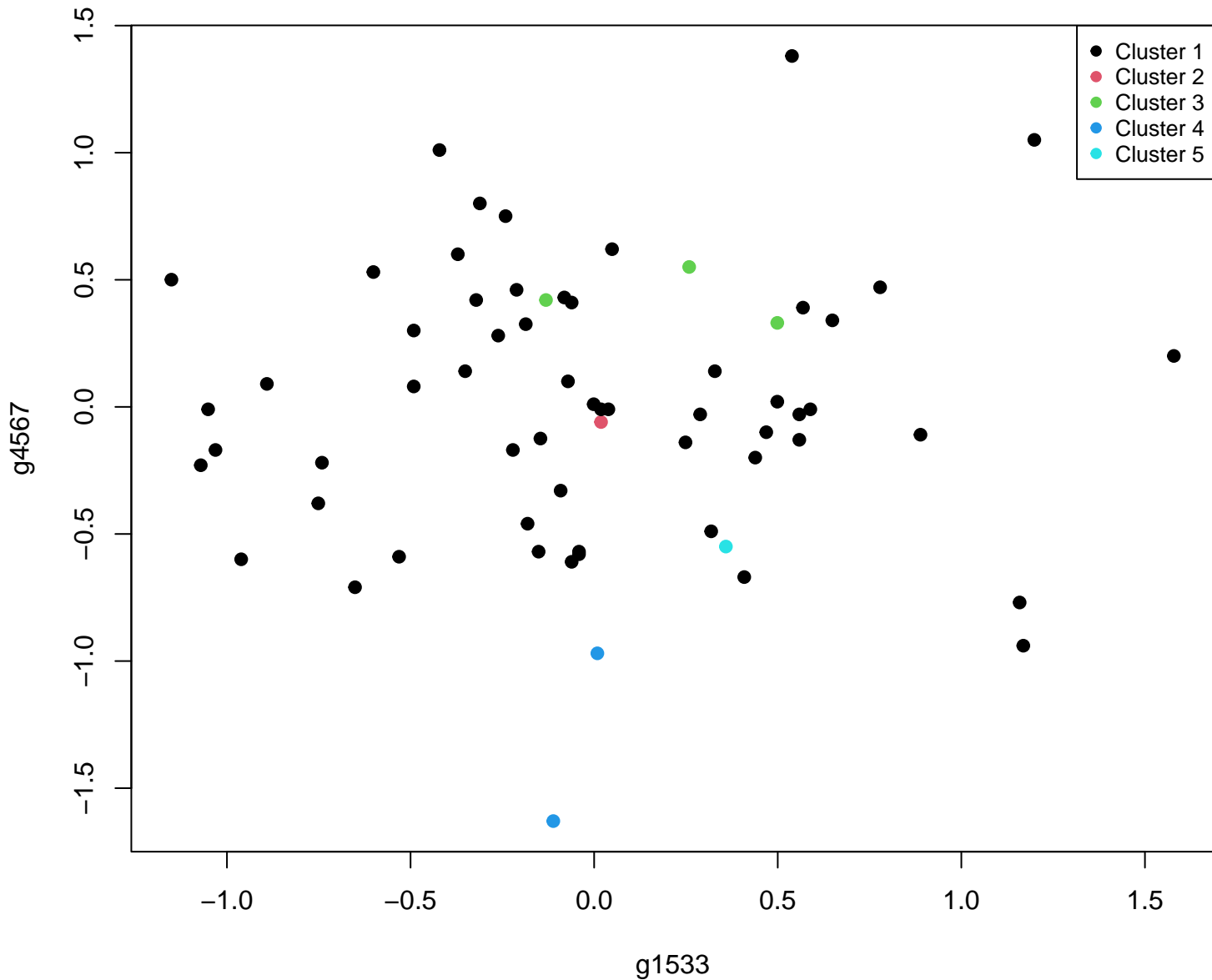




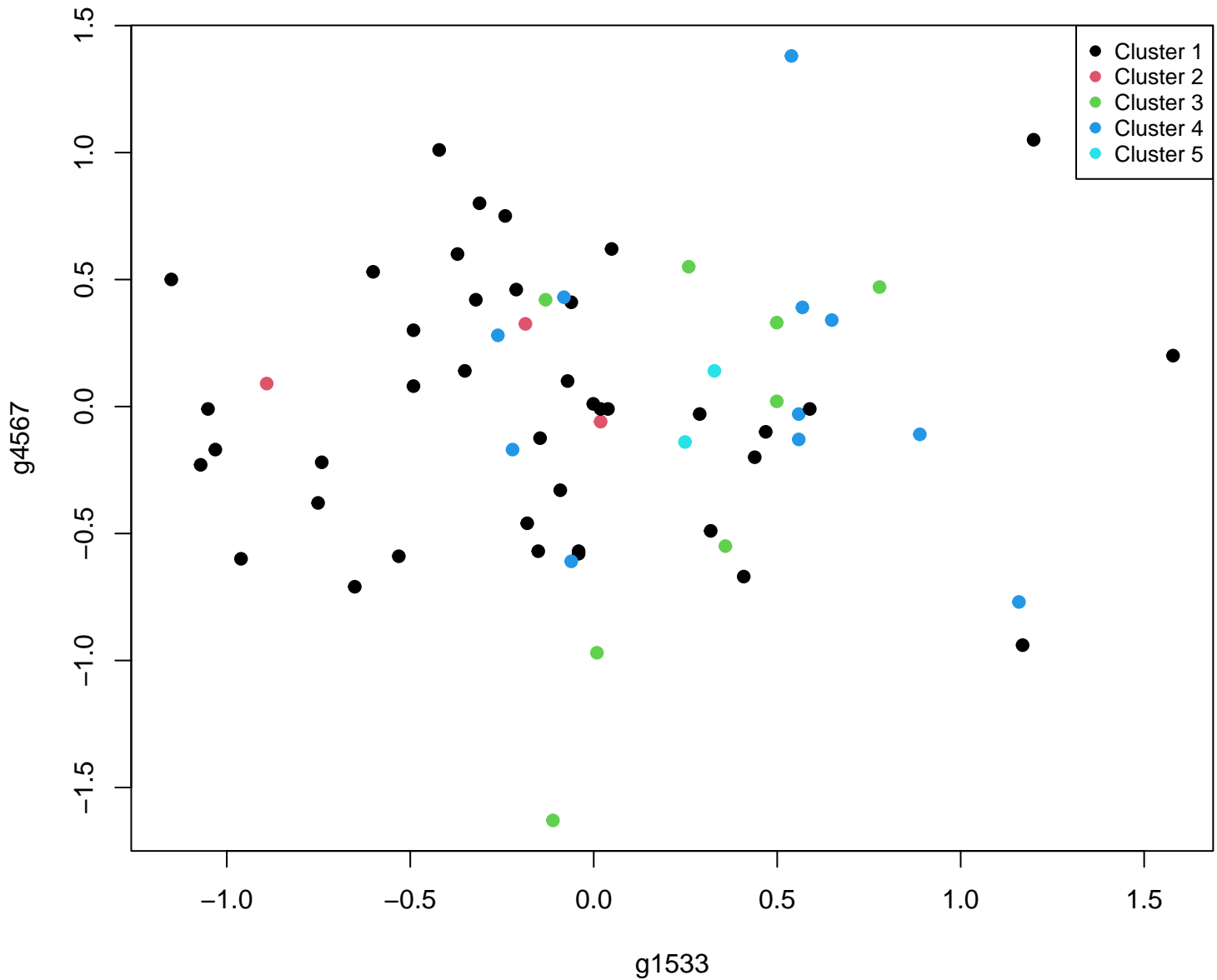
**Average linkage clusters with 5 cluster cutoff (using 2 randomly sampled genes)**



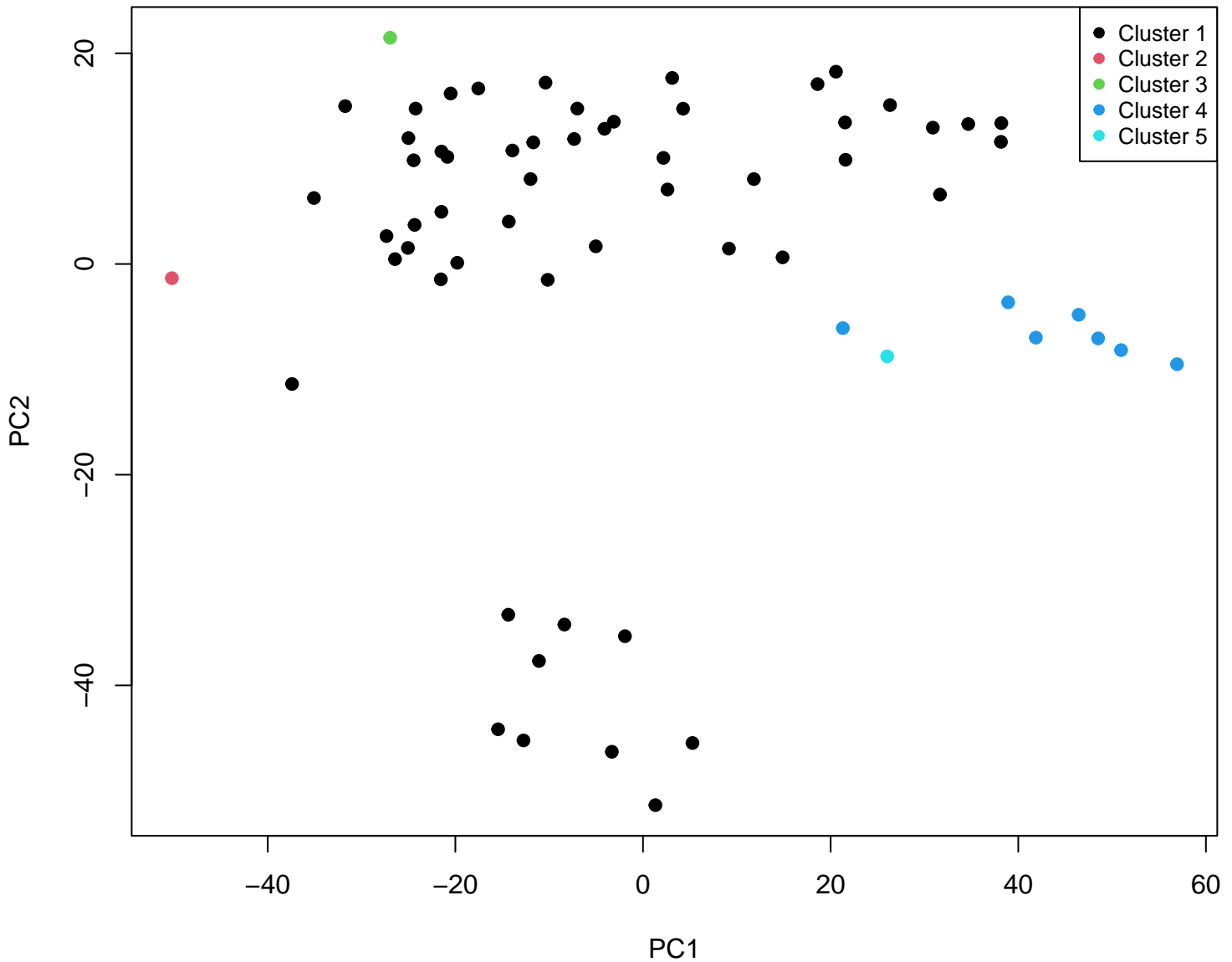
Single linkage clusters with 5 cluster cutoff (using 2 randomly sampled genes)



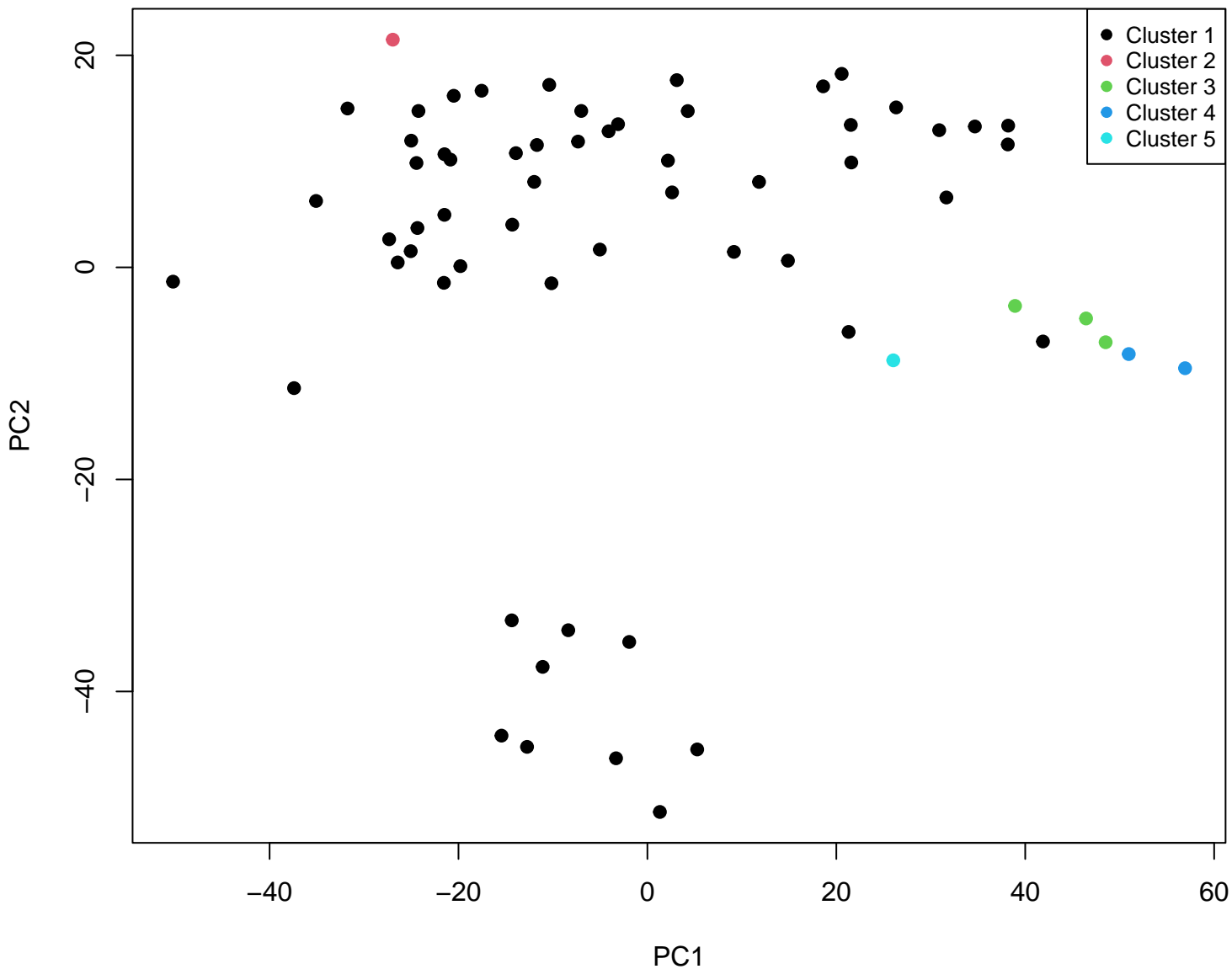
Complete linkage clusters with 5 cluster cutoff (using 2 randomly sampled genes)



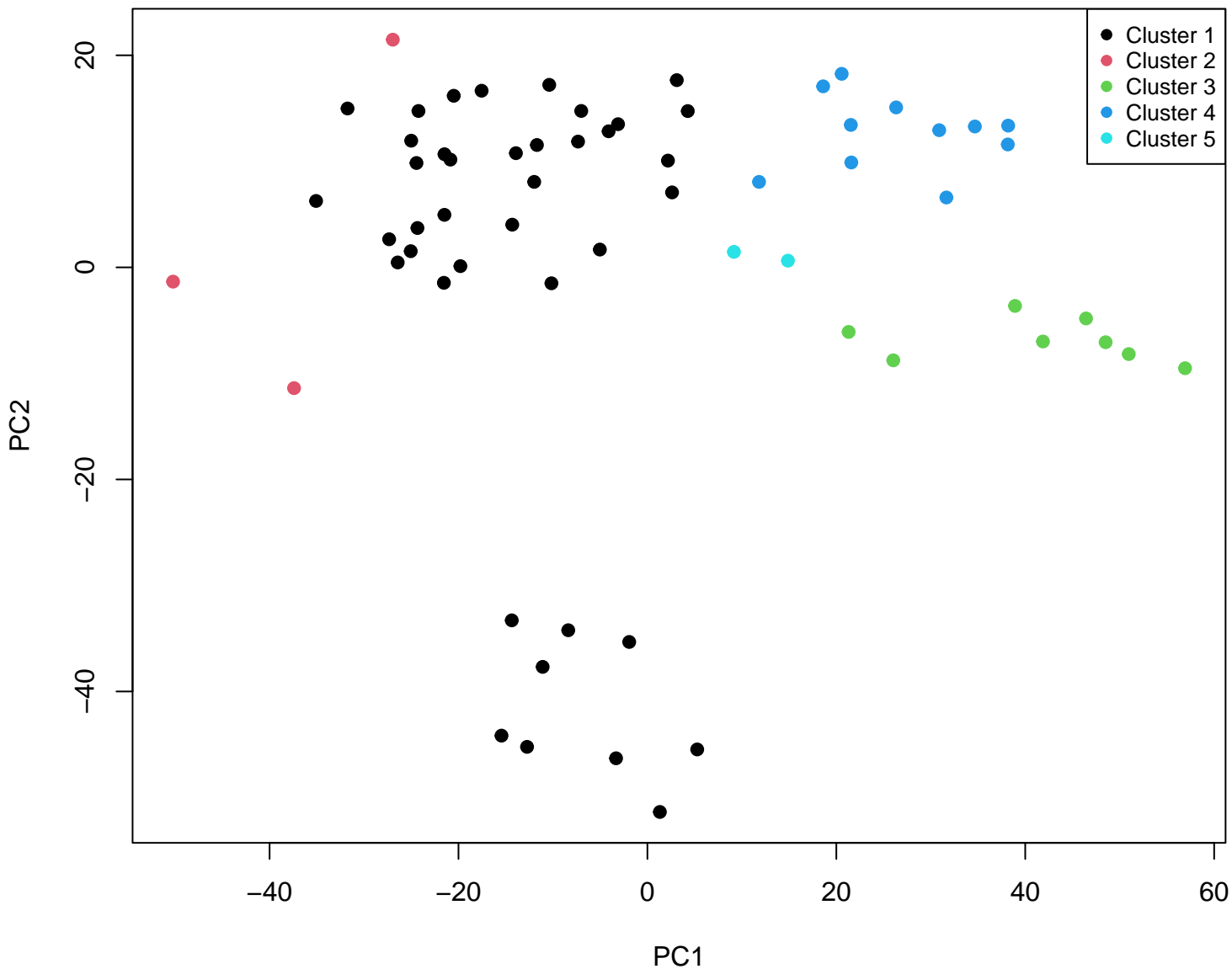
Average linkage clusters (first 2 PCs, 5 clusters cut-off)



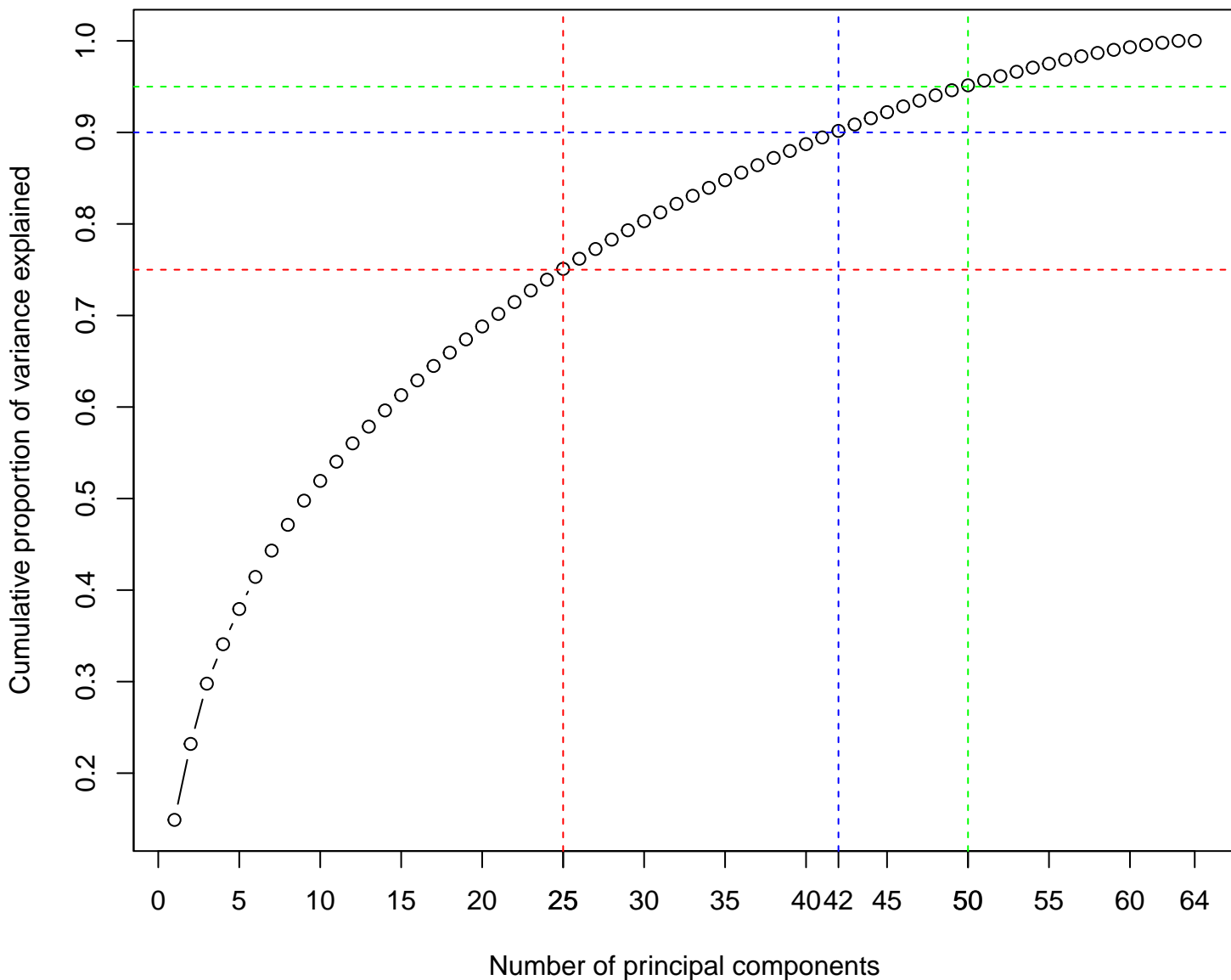
Single linkage clusters (first 2 PCs, 5 clusters cut-off)



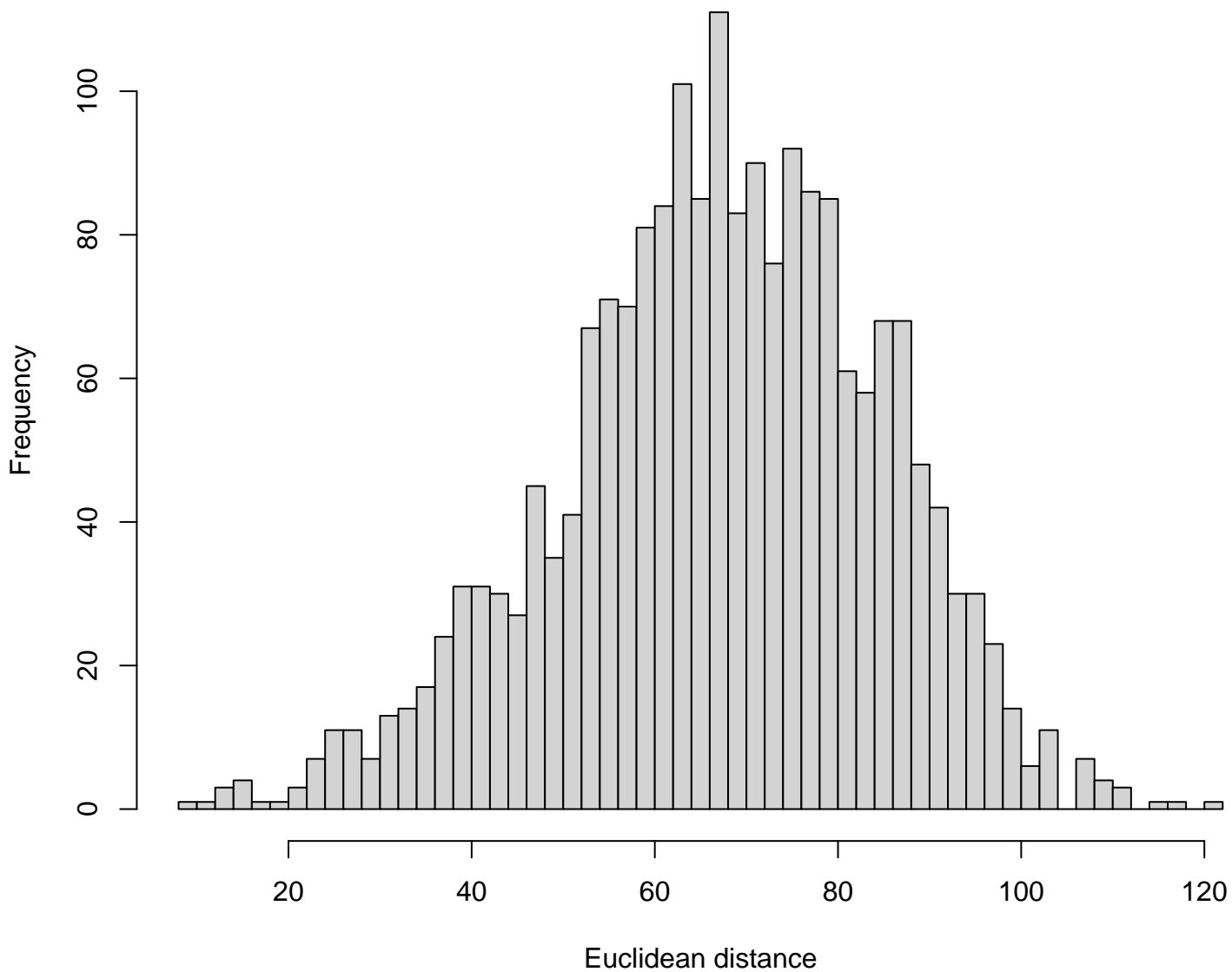
**Complete linkage clusters (first 2 PCs, 5 clusters cut-off)**



**Cumulative variance explained by PCA**

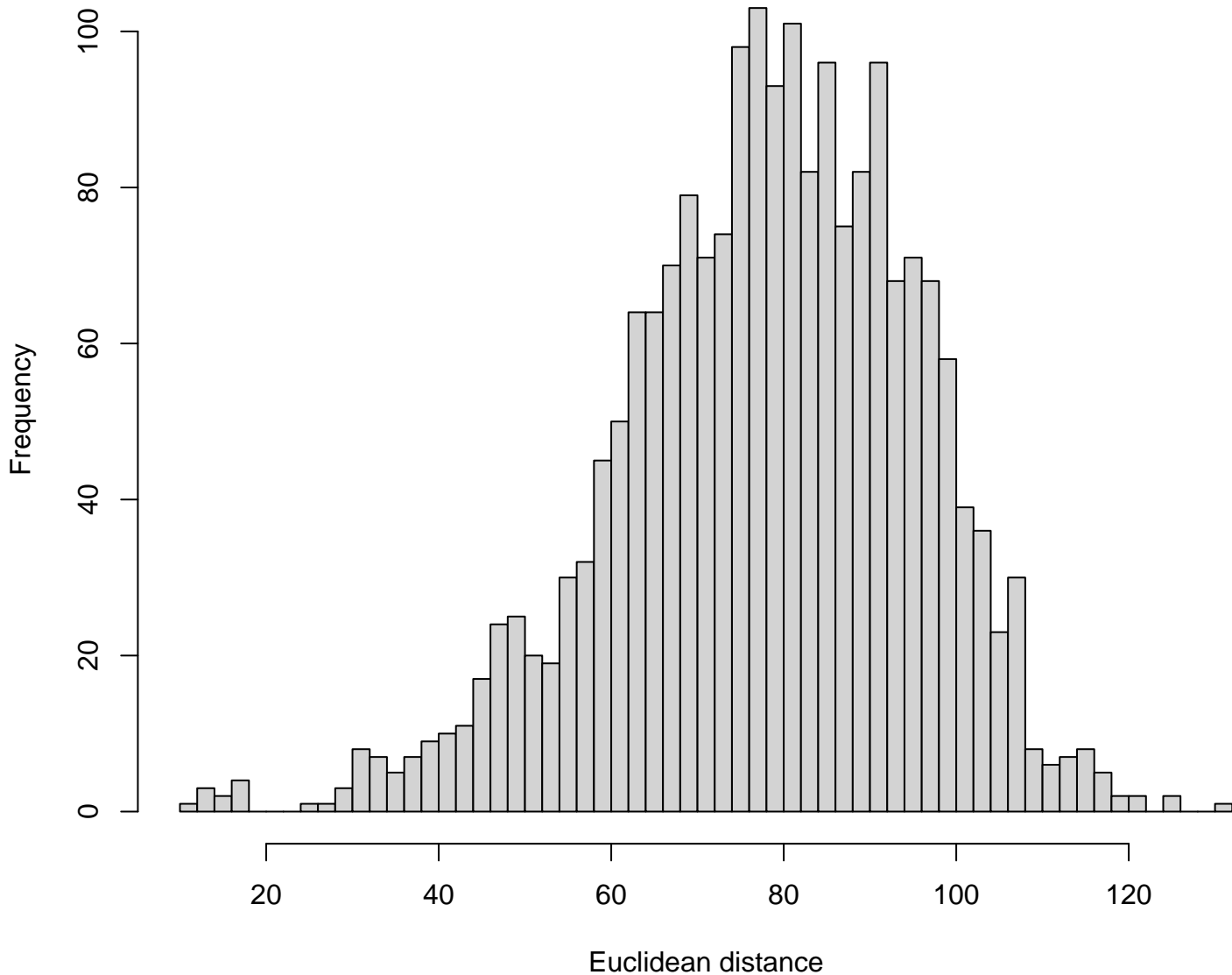


# Pairwise distances between samples on reduced data Z\_sp

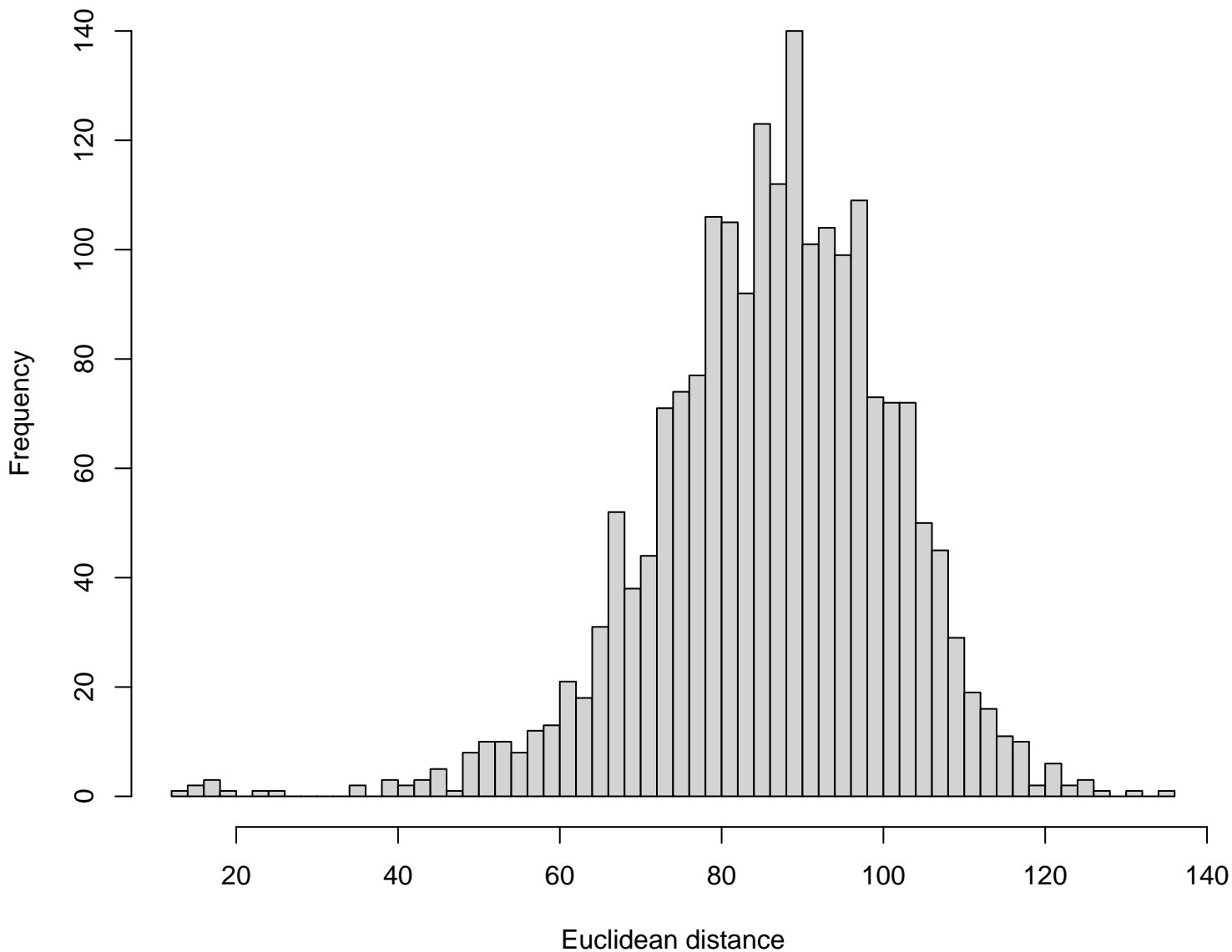




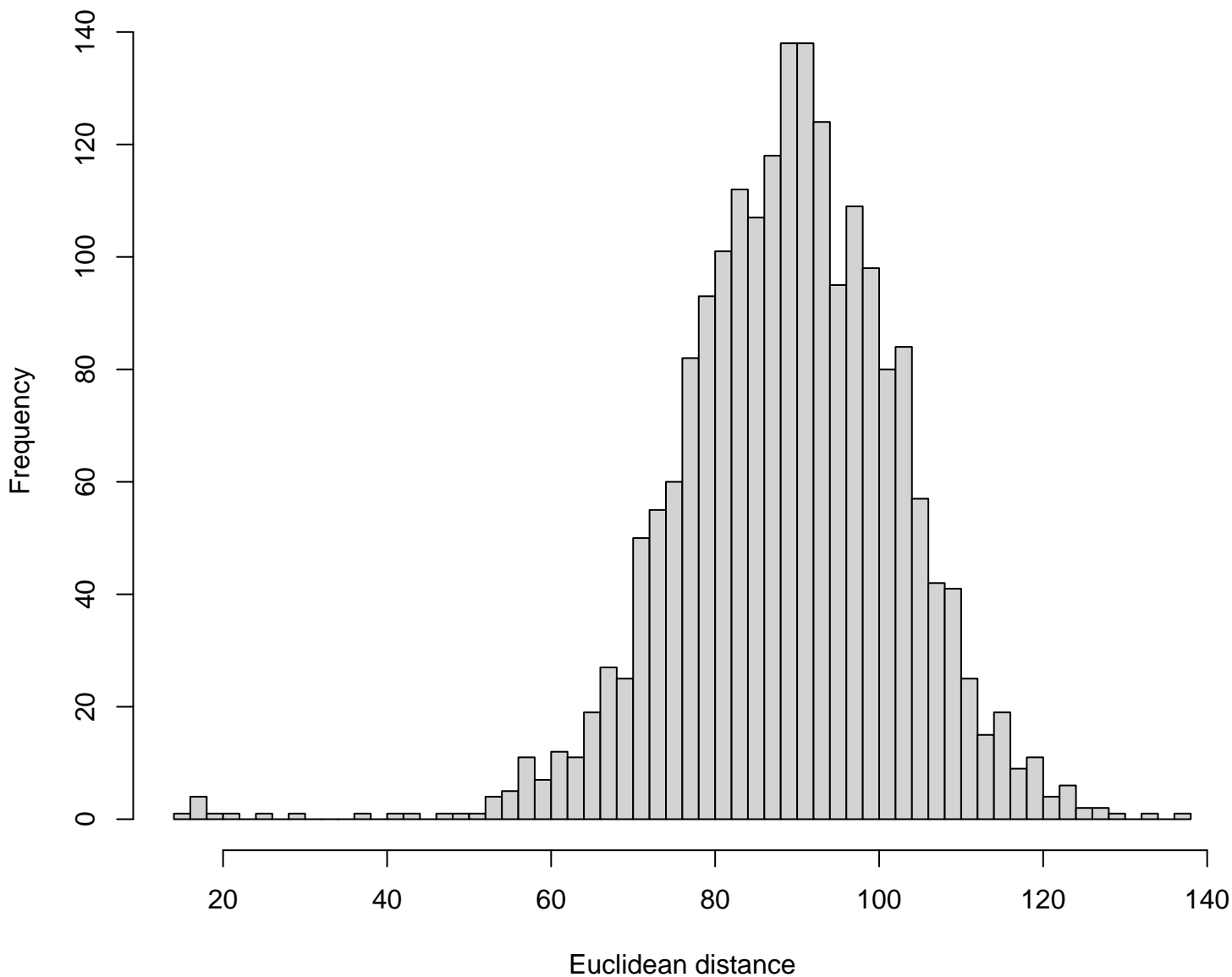
# Pairwise distances between samples on reduced data with Z\_75



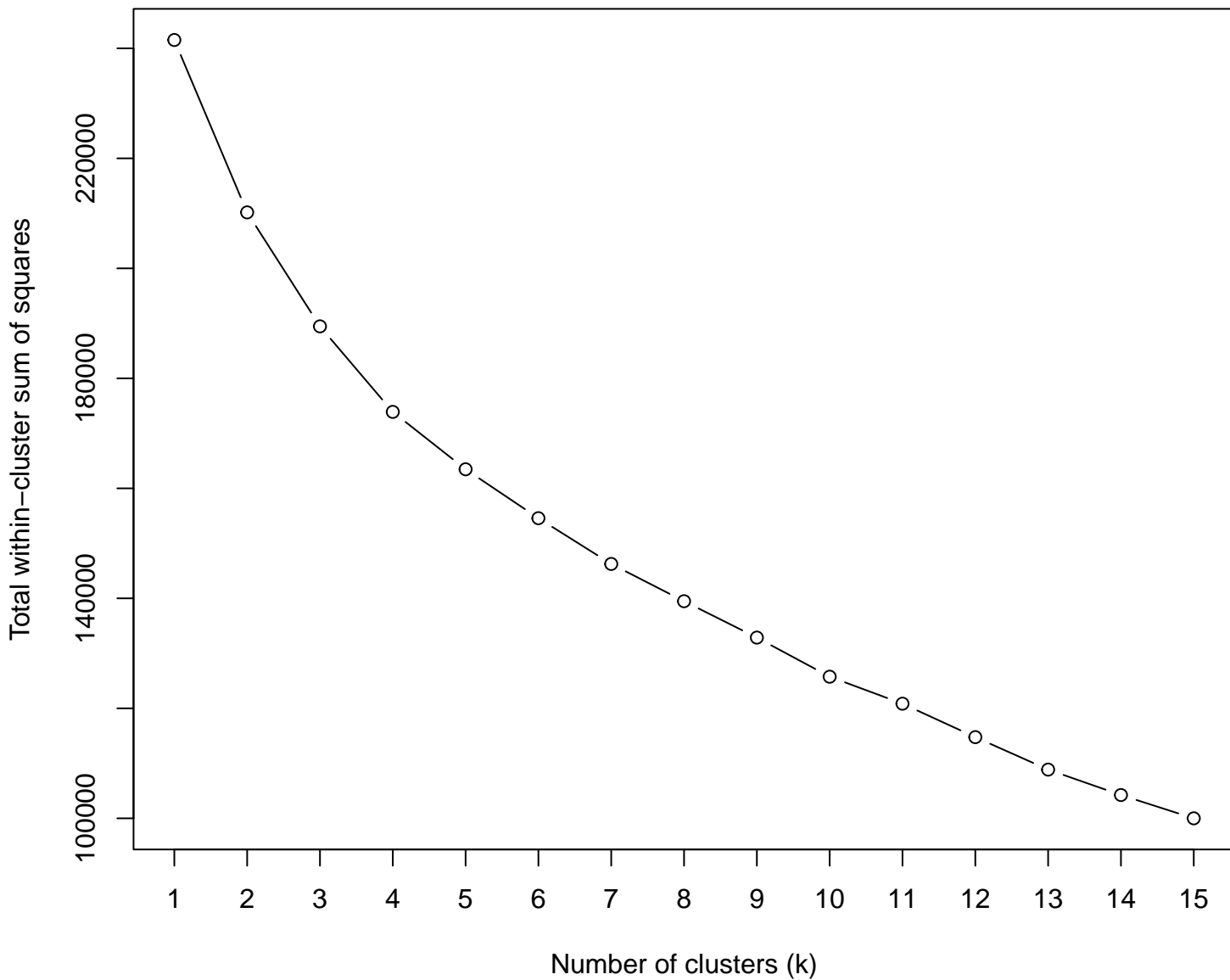
# Pairwise distances between samples on reduced data with Z\_90



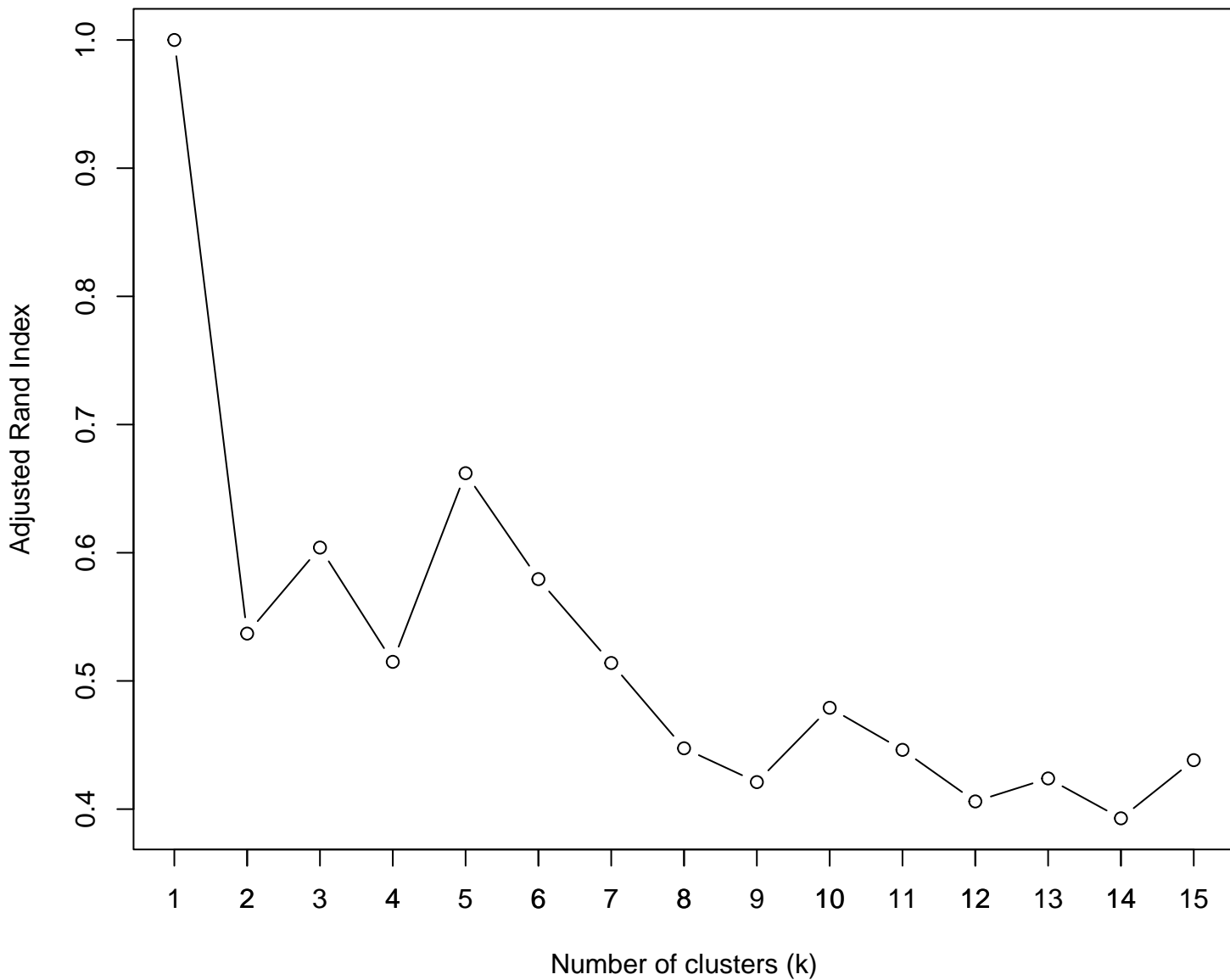
# Pairwise distances between samples on reduced data with Z\_95



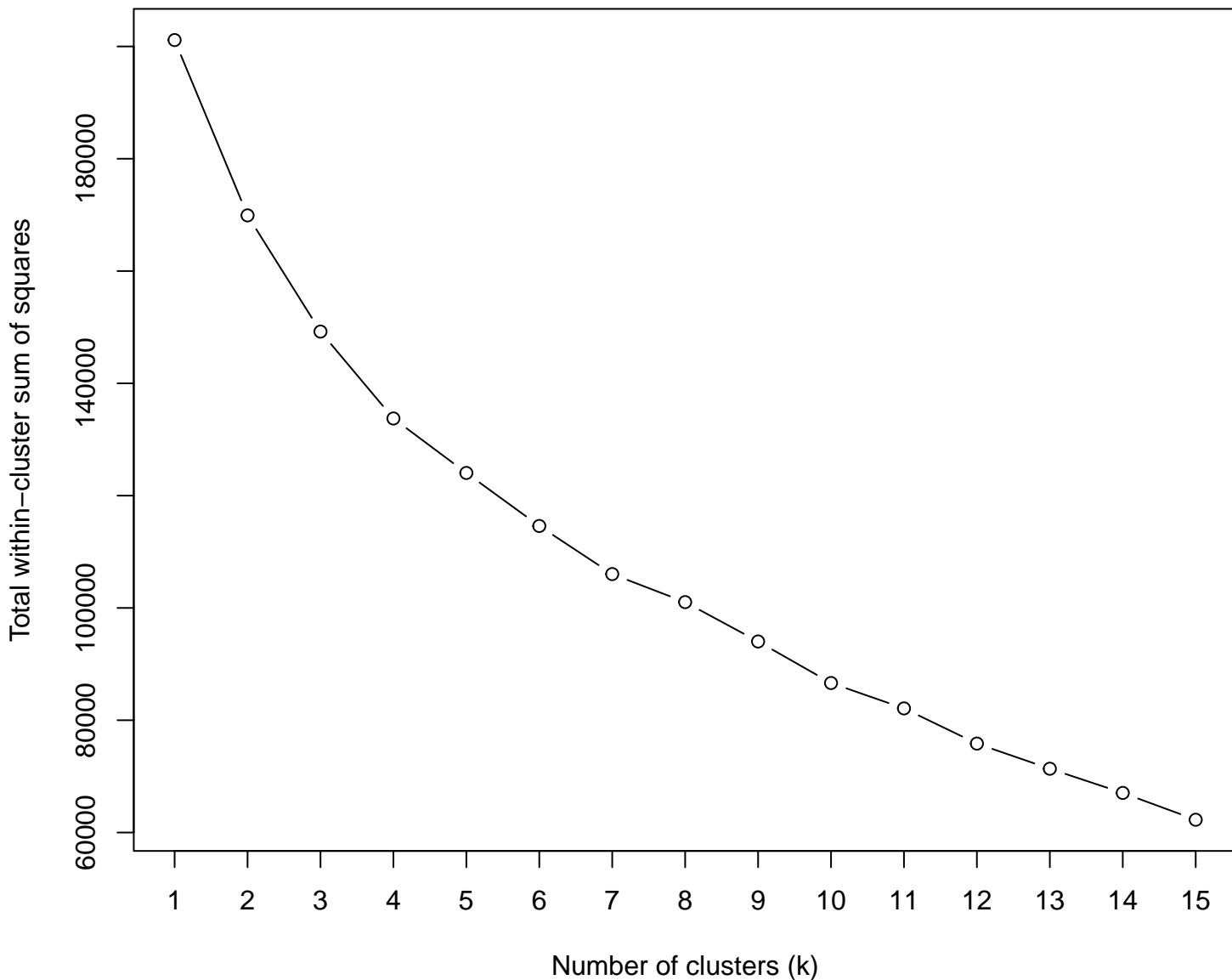
**Elbow plot for k-means clustering on reduced data Z**



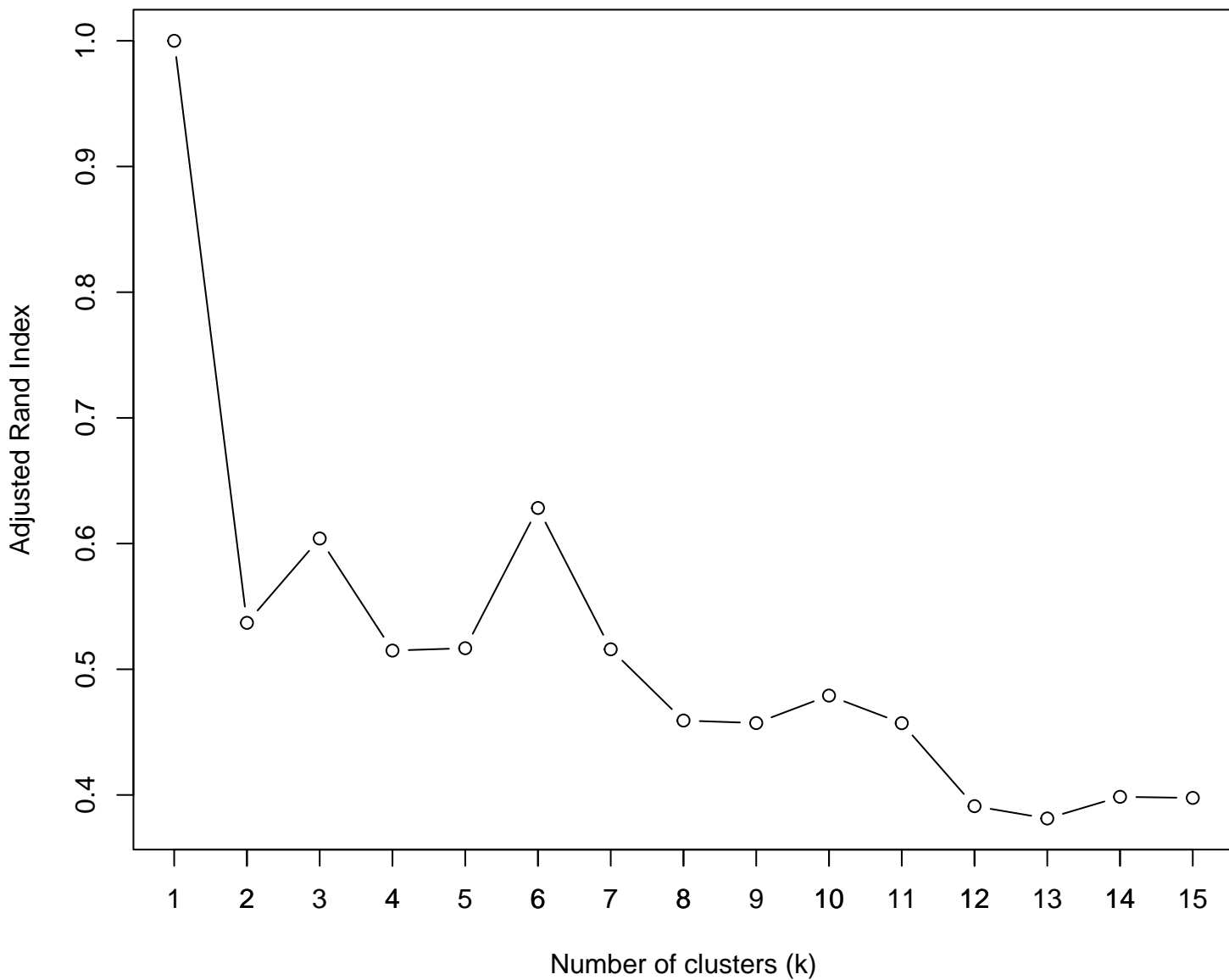
**ARI vs k clusters on reduced data Z**



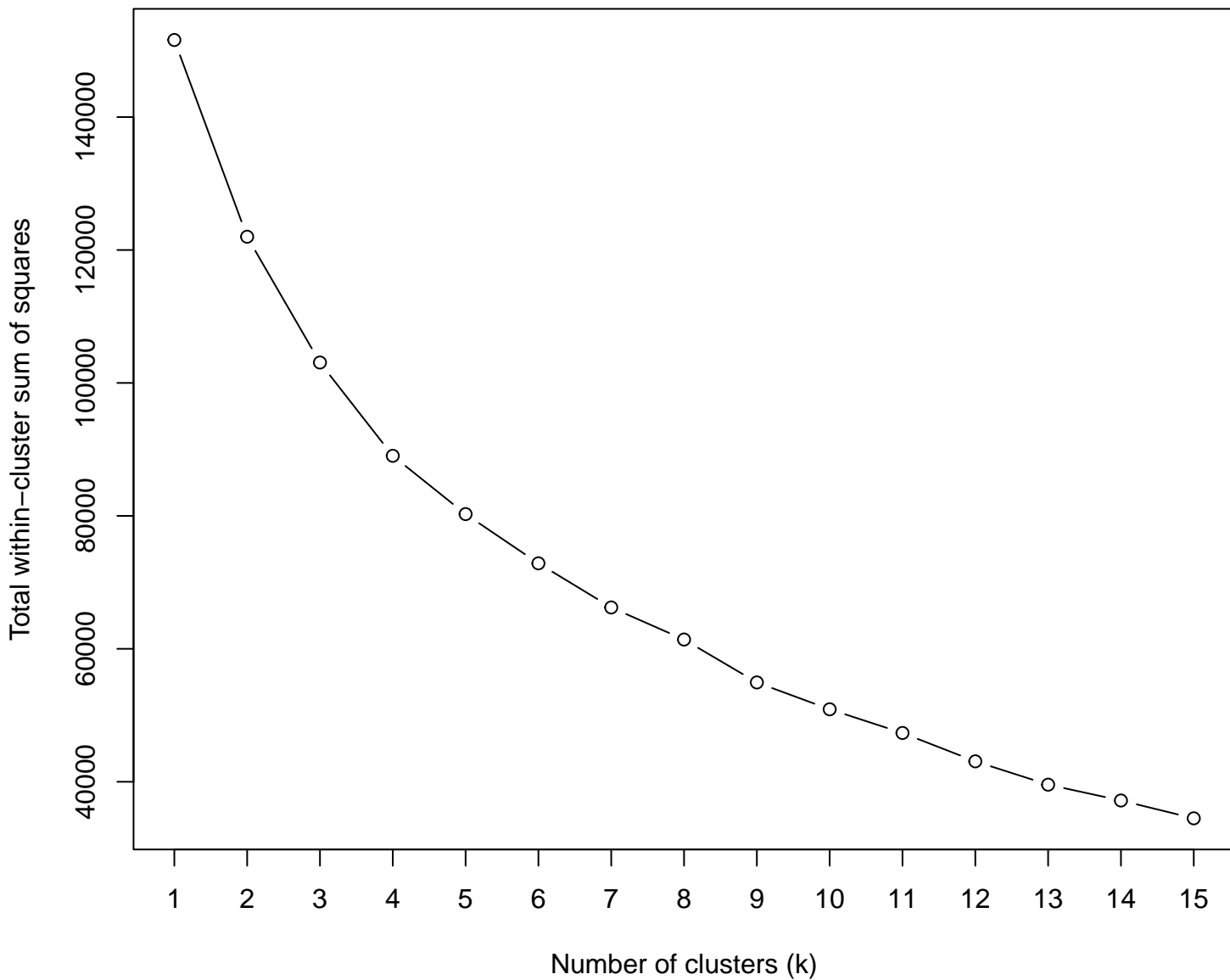
**Elbow plot for k-means clustering on reduced data Z\_75**



**ARI vs k clusters on reduced data Z\_75**

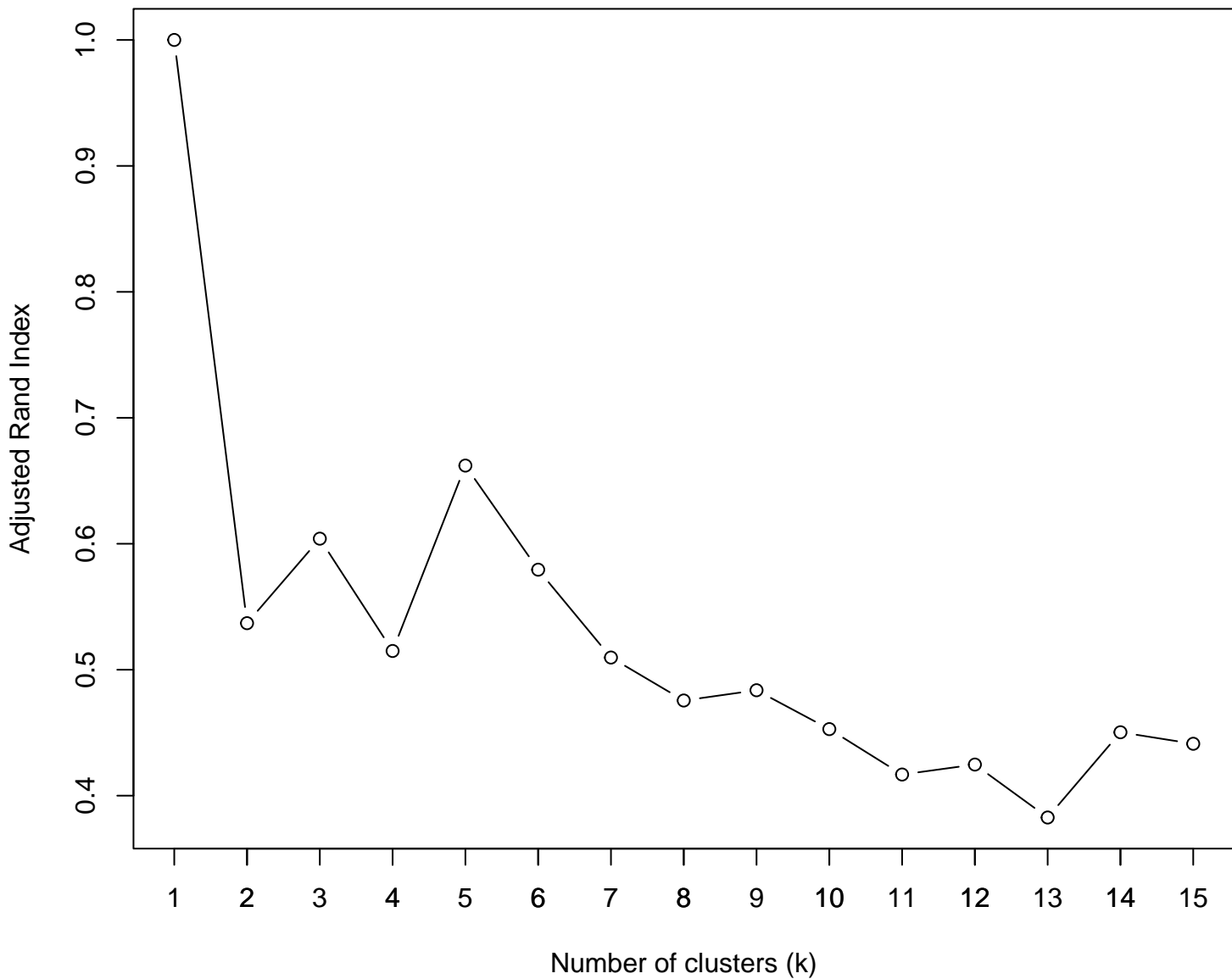


**Elbow plot for k-means clustering on reduced data Z\_sp**

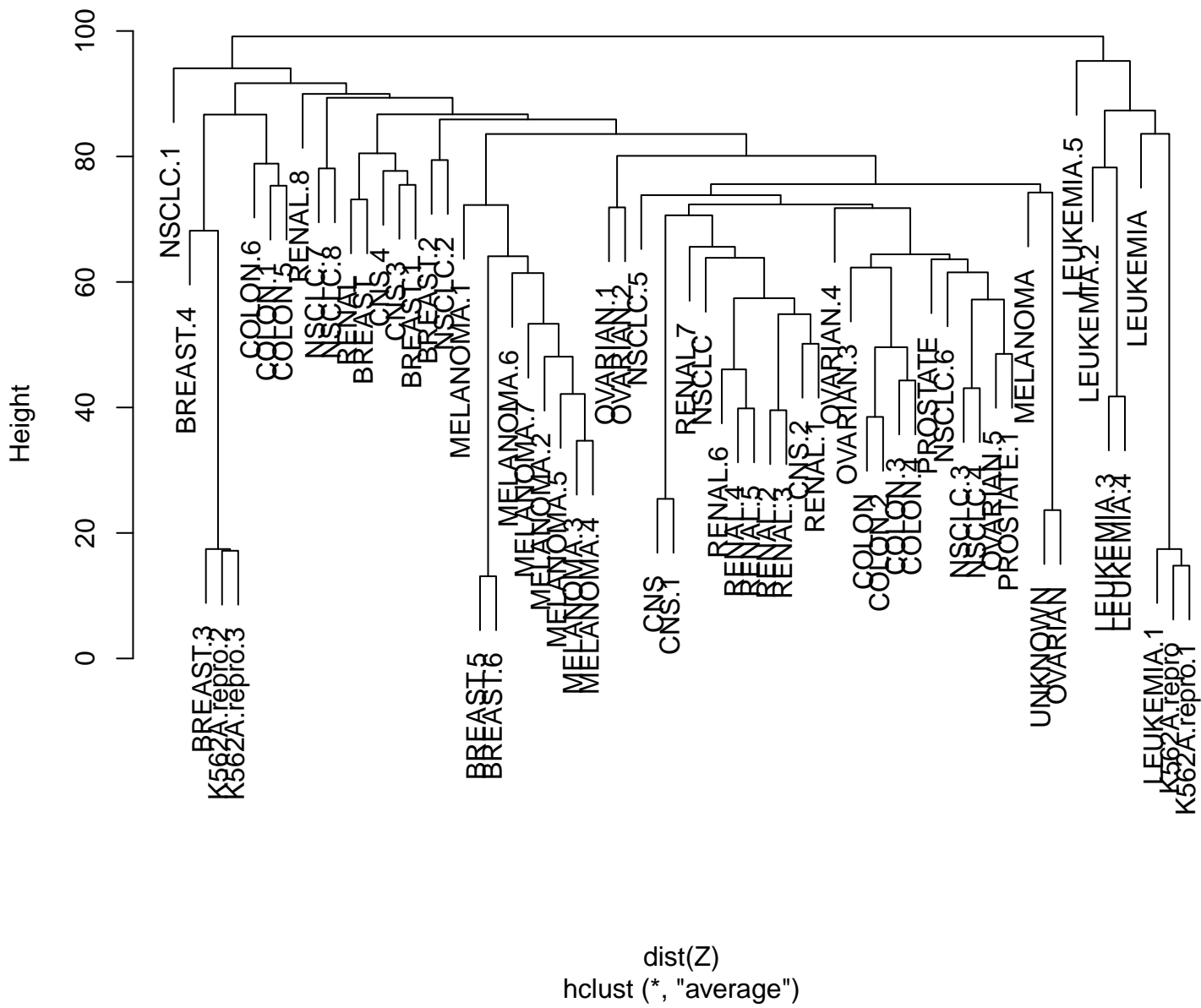




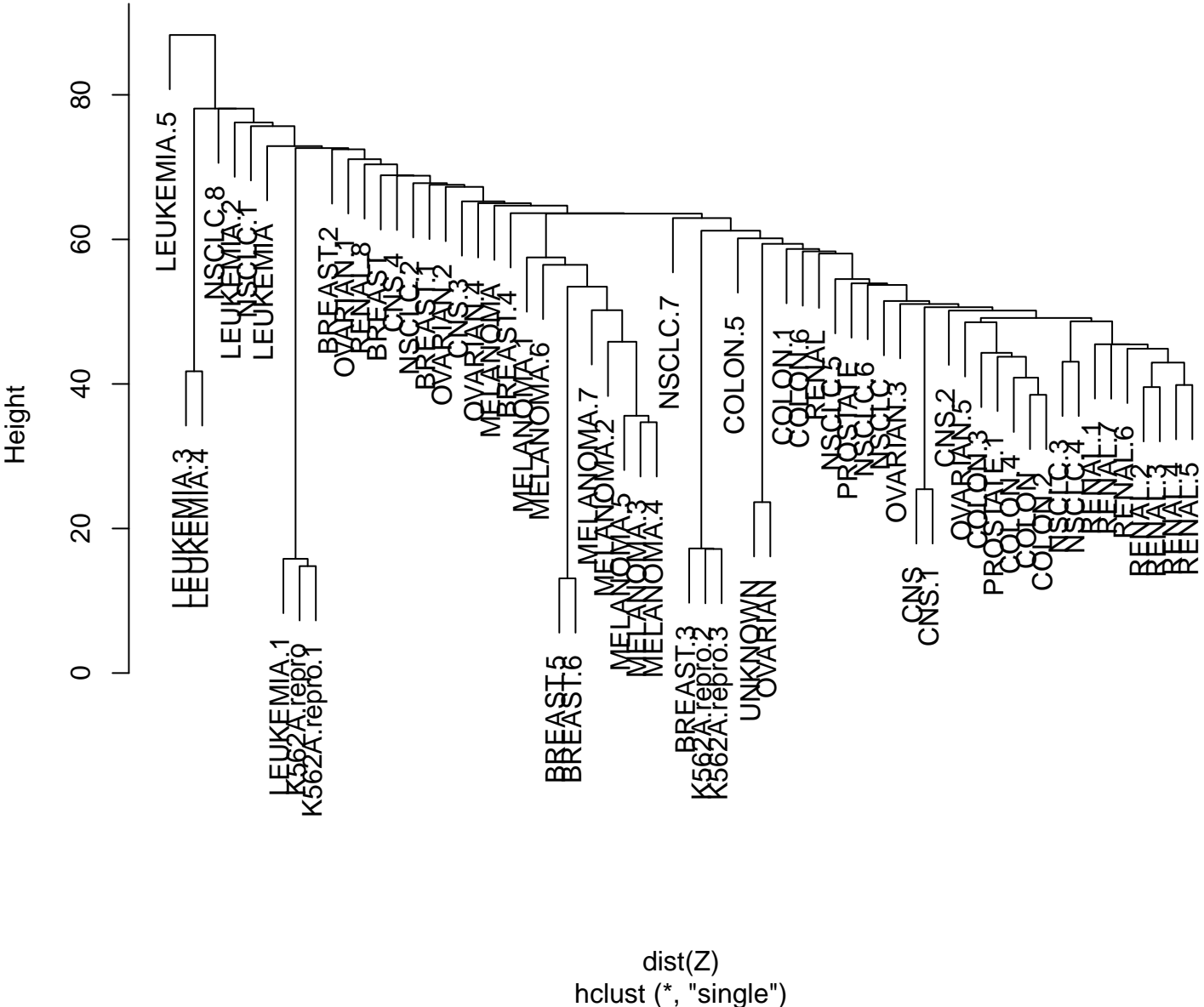
**ARI vs k clusters on reduced data Z\_sp**



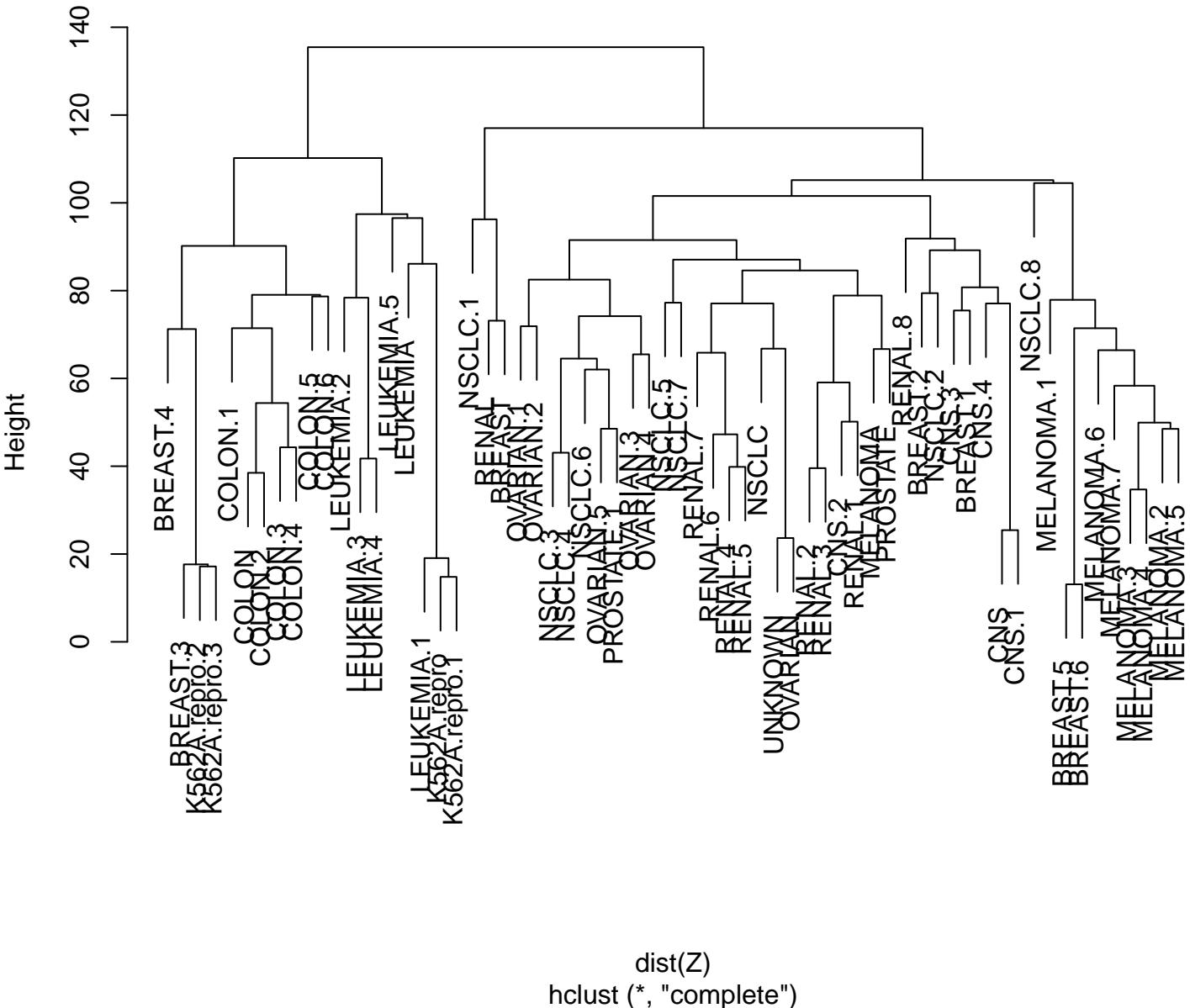
# Average linkage on reduced data Z



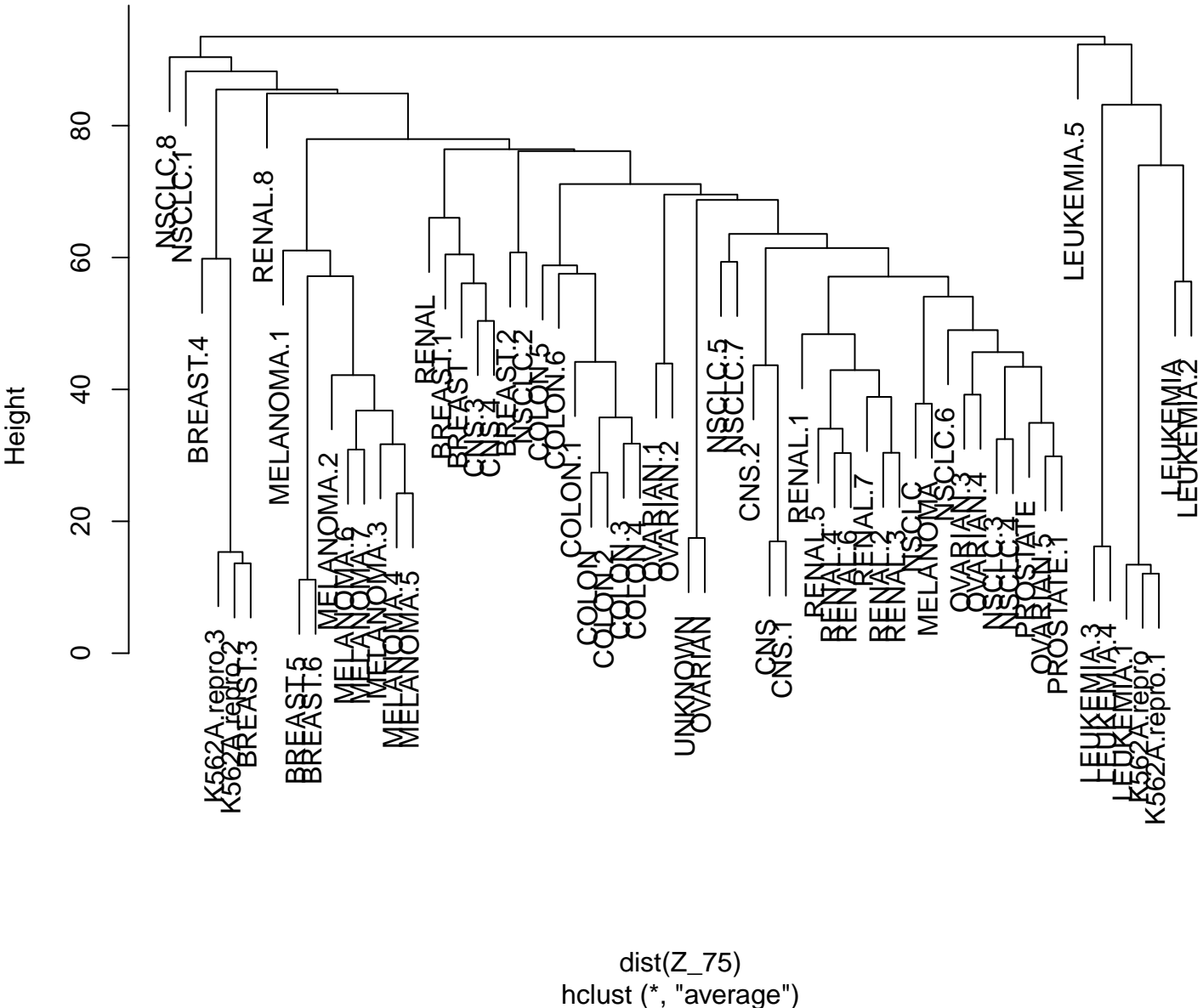
## Single linkage on reduced data Z



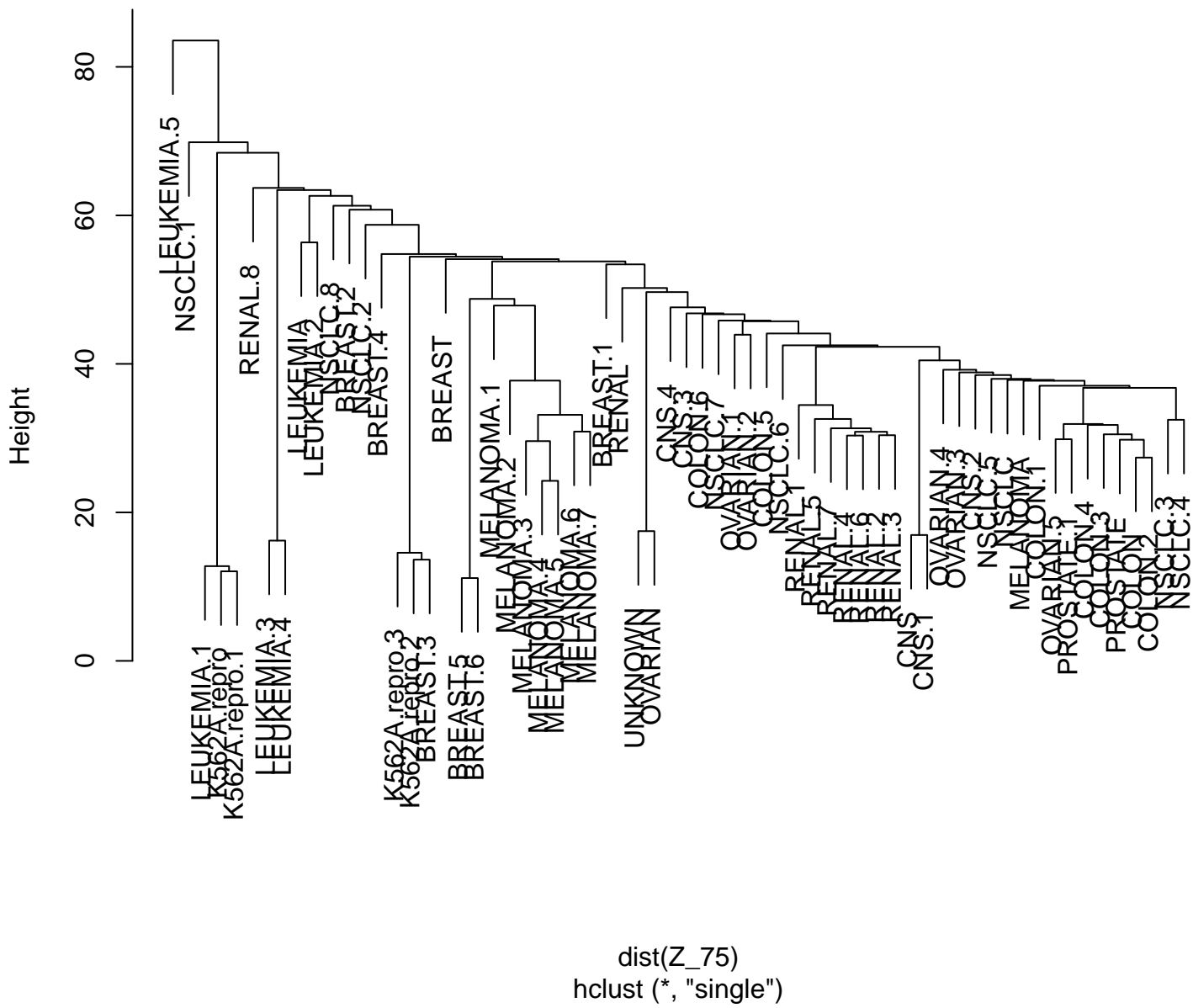
## Complete linkage on reduced data Z



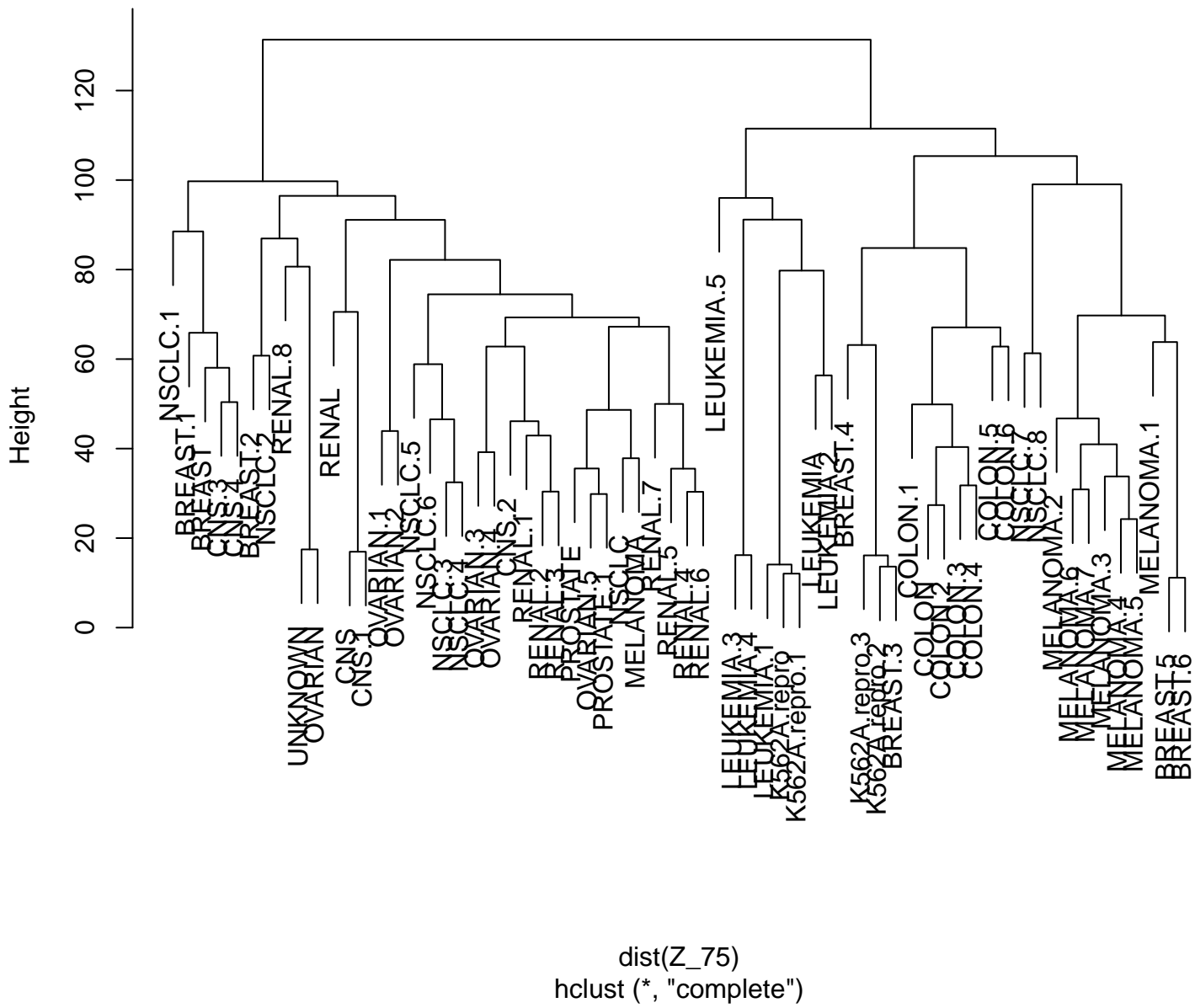
### Average linkage on reduced data Z\_75



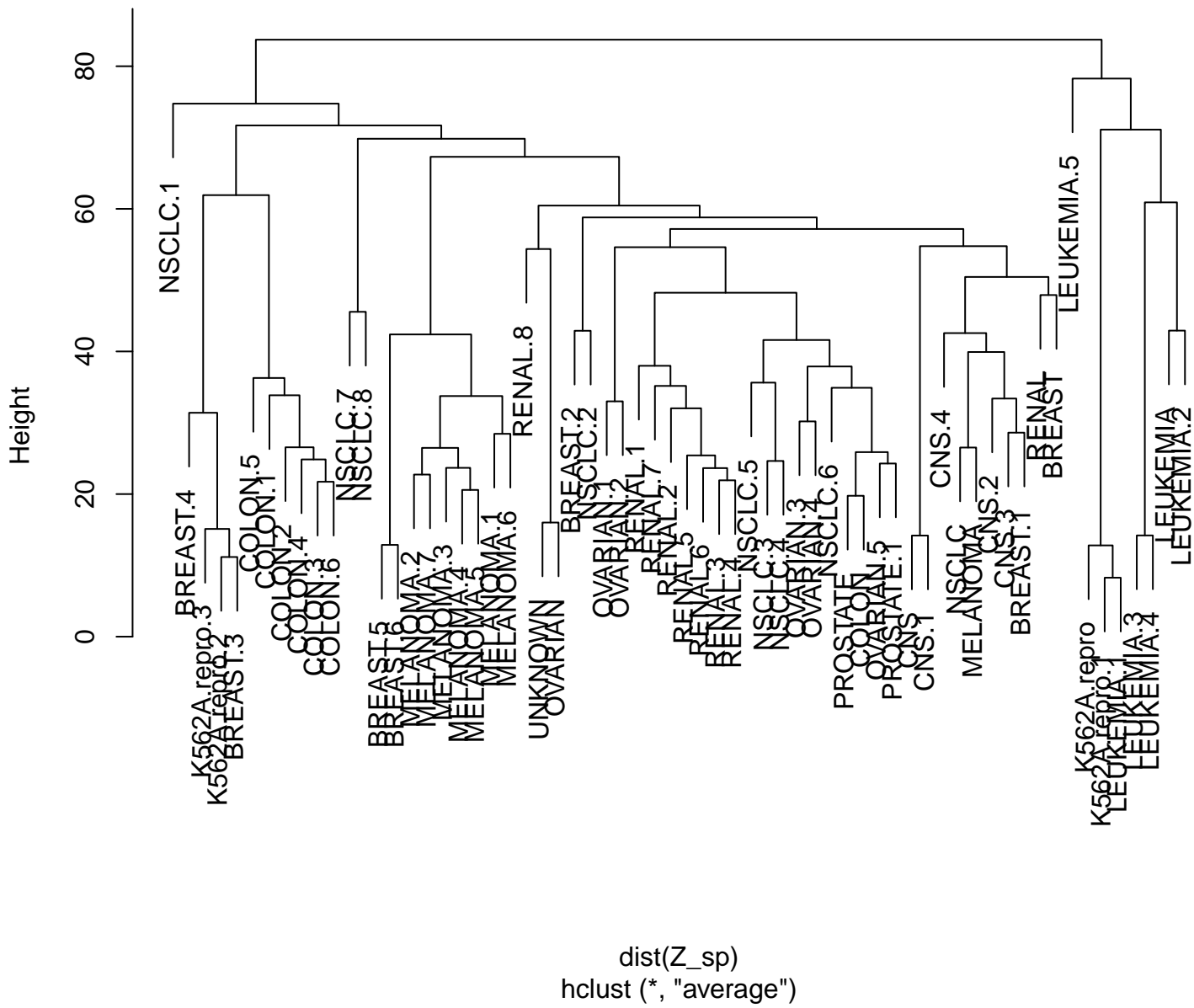
# Single linkage on reduced data Z\_75



# Complete linkage on reduced data Z\_75

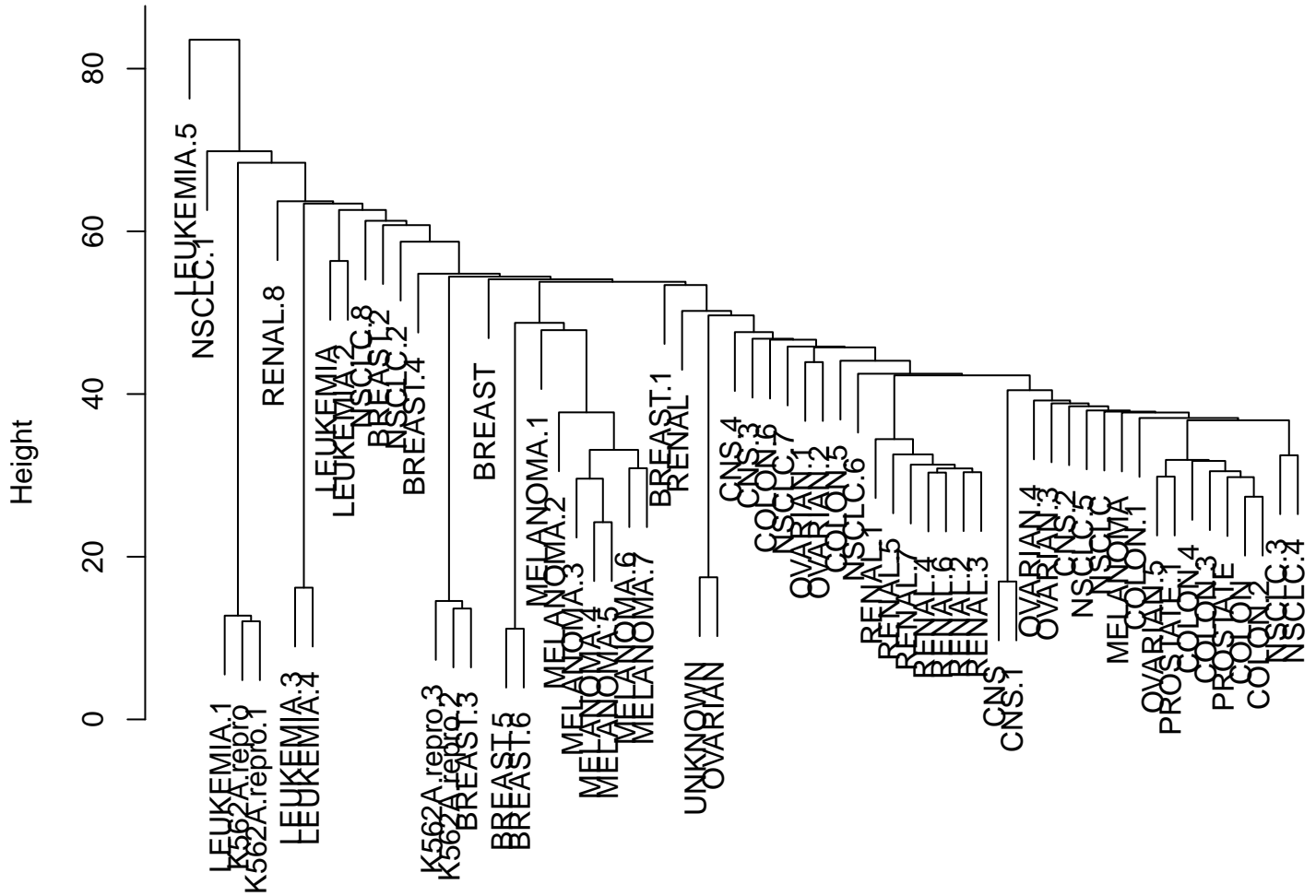


### Average linkage on reduced data Z\_sp

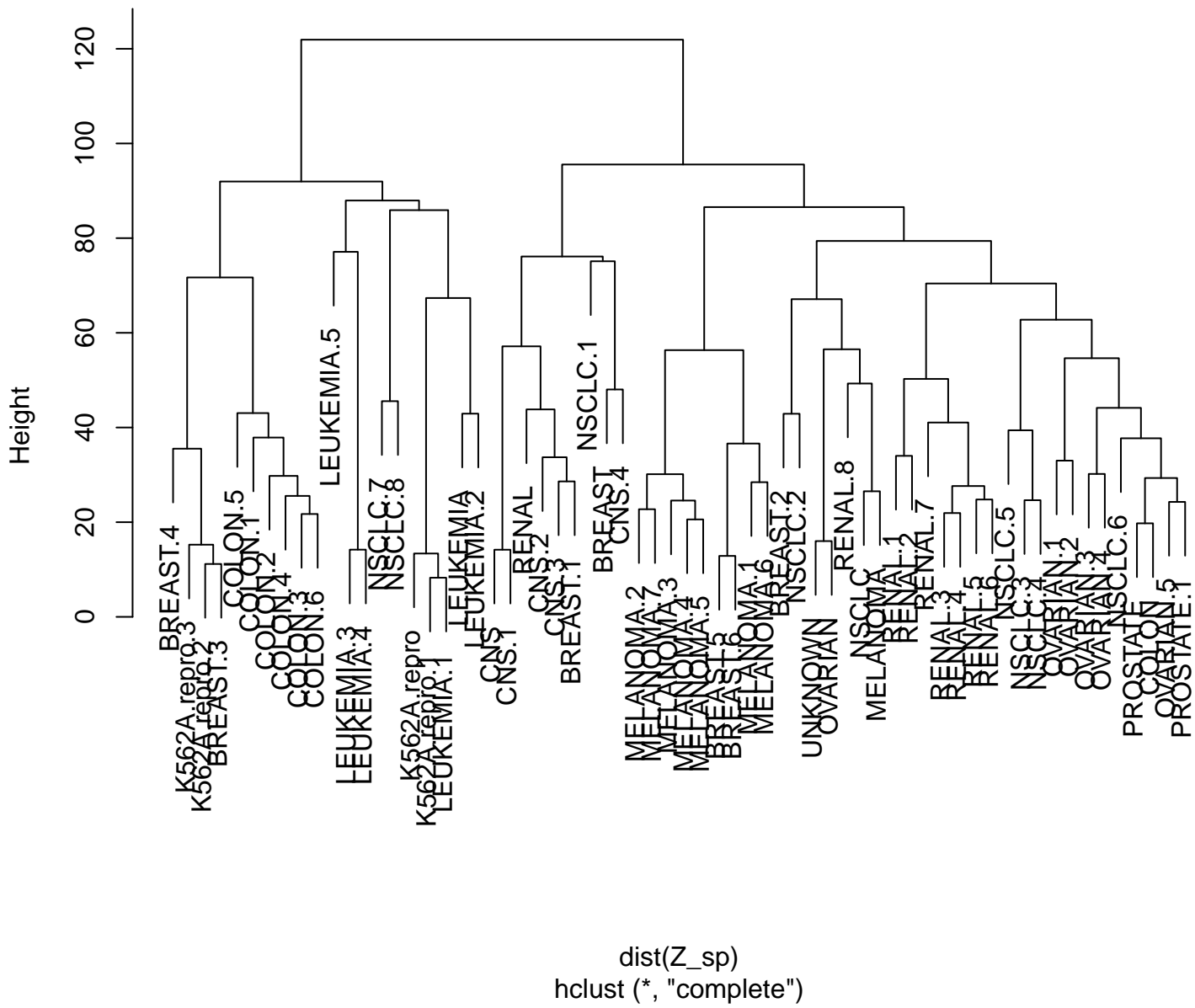




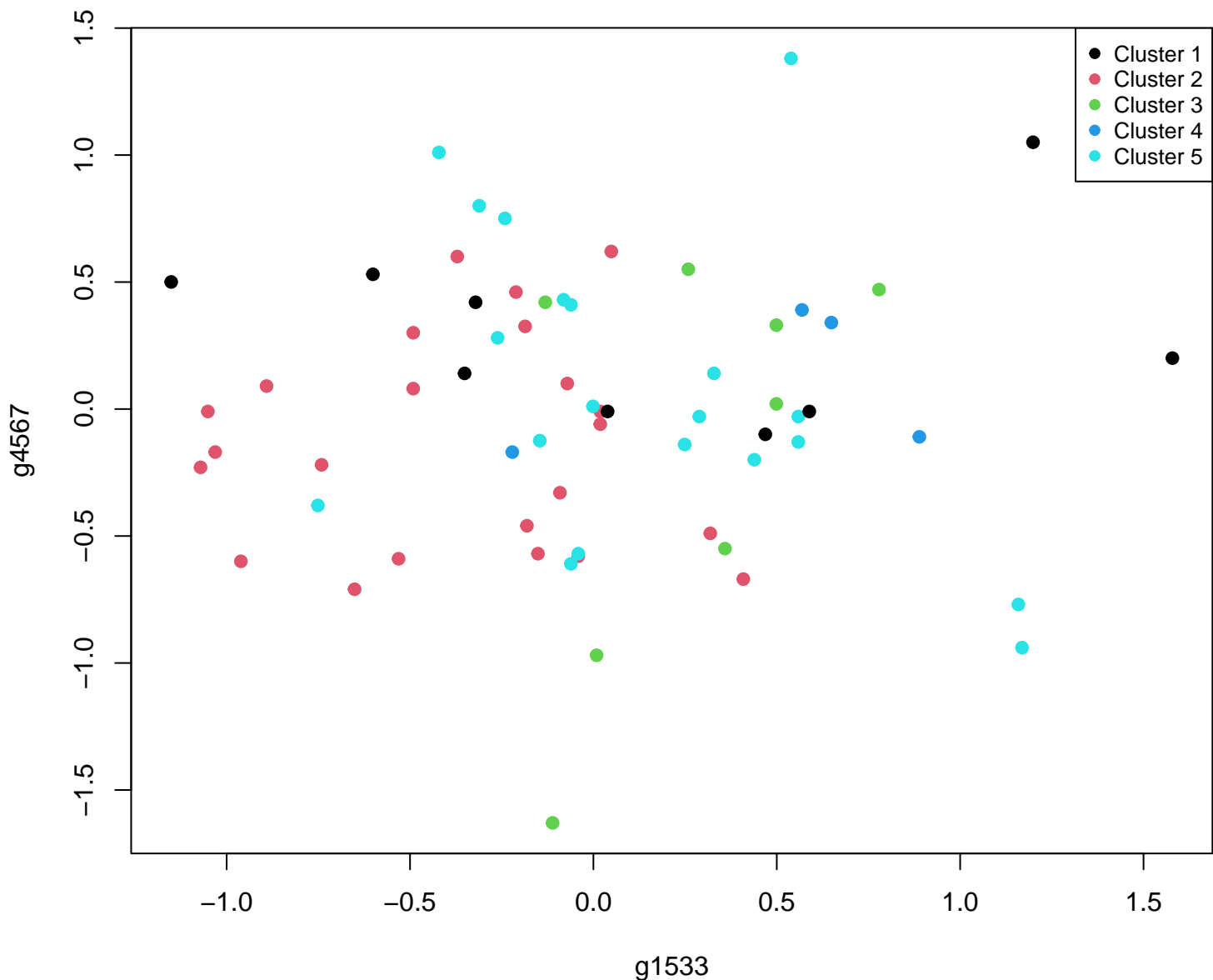
```
dist(Z_75)
hclust (*, "single")
```



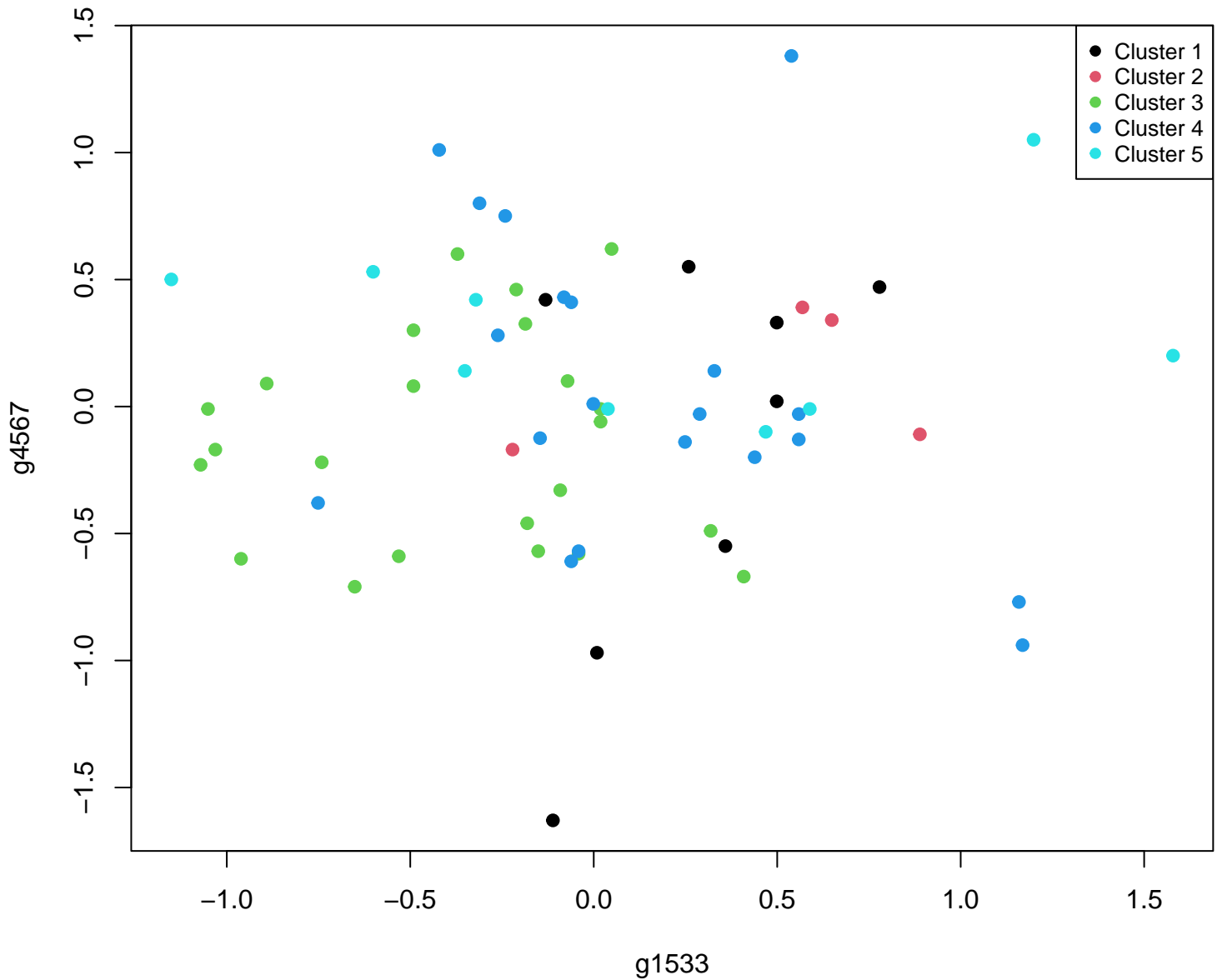
# Complete linkage on reduced data Z\_sp



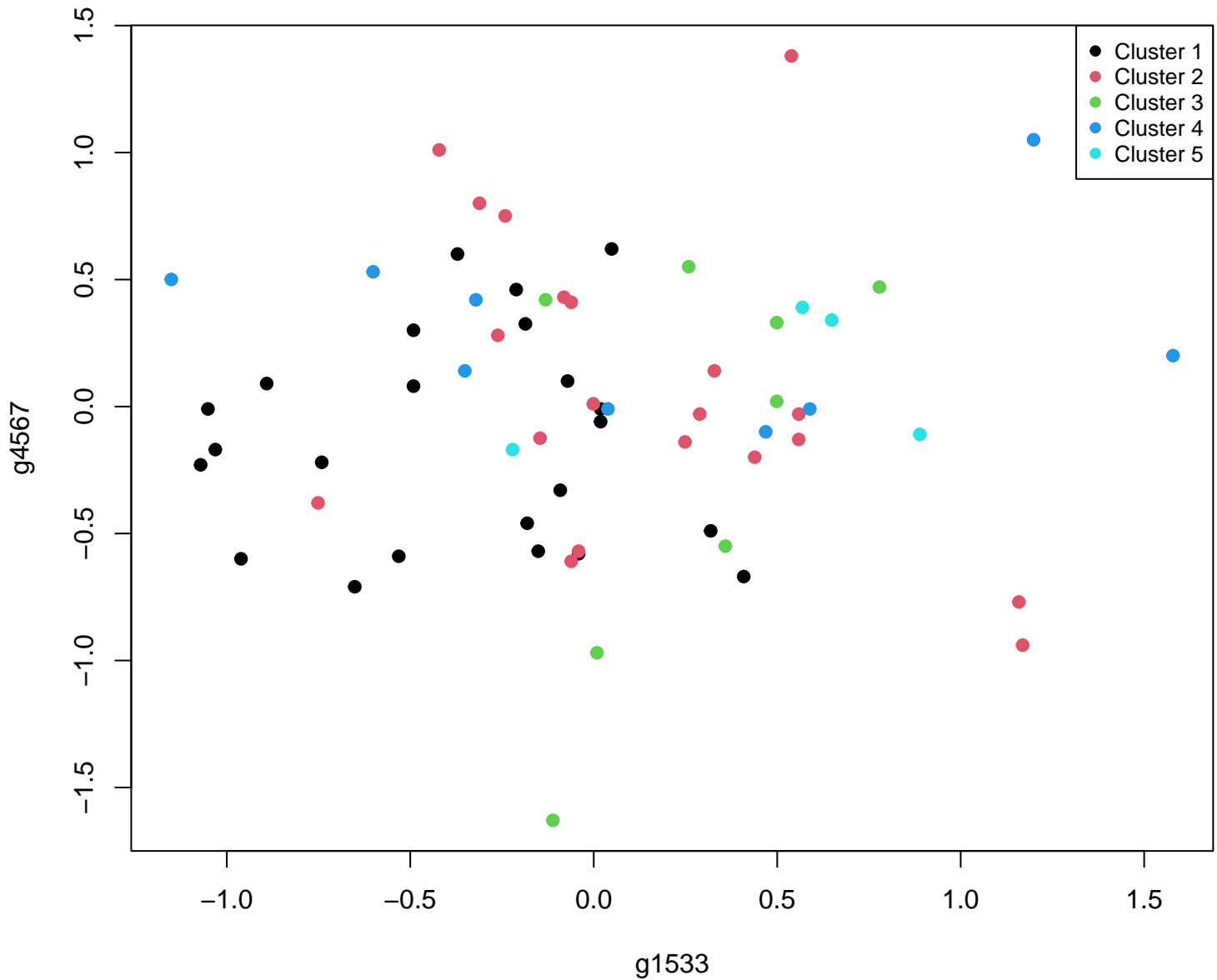
Reduced data Z k-means clusters with 5 clusters (using 2 randomly sampled genes



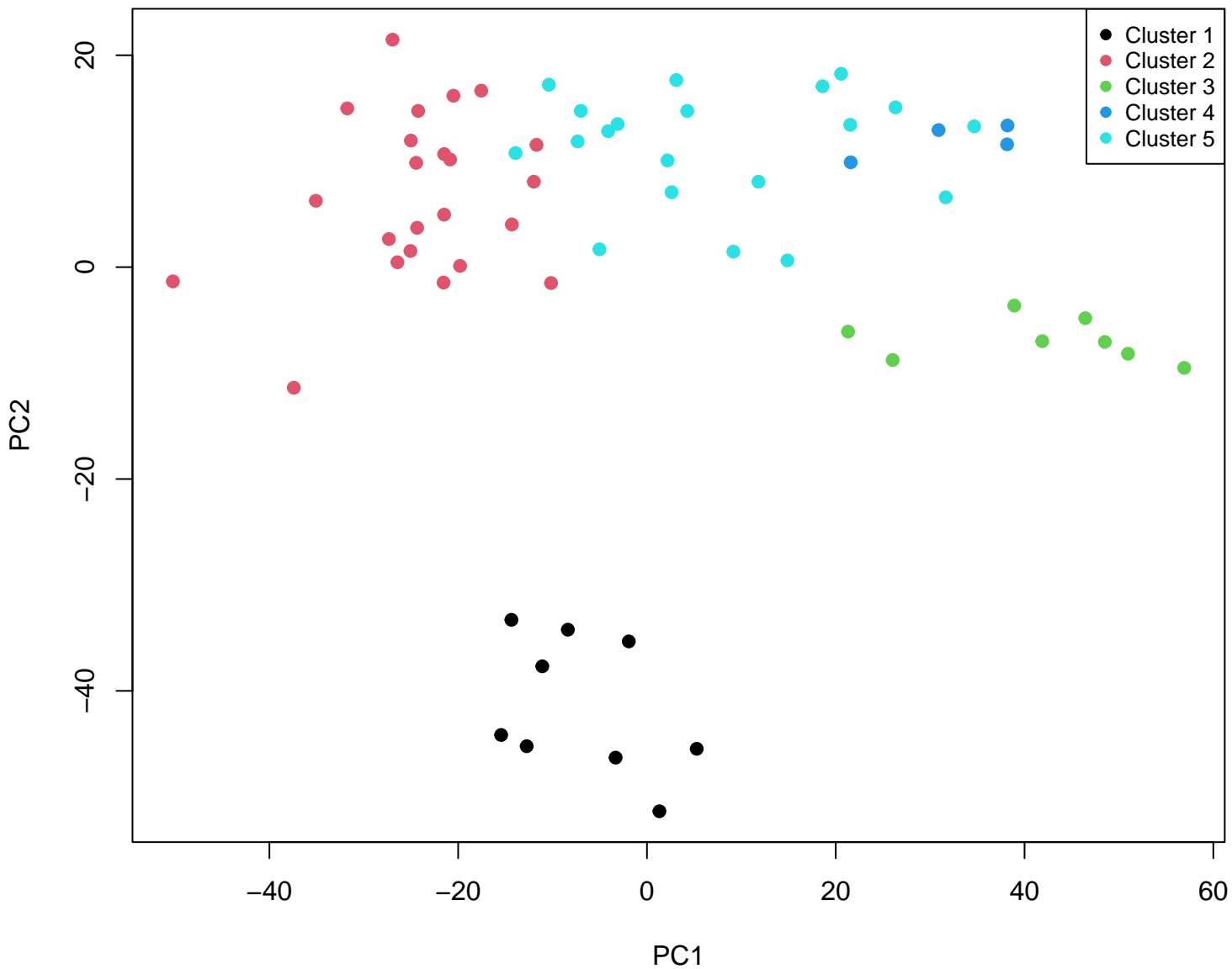
Reduced data Z\_75 k-means clusters with 5 clusters (using 2 randomly sampled genes)



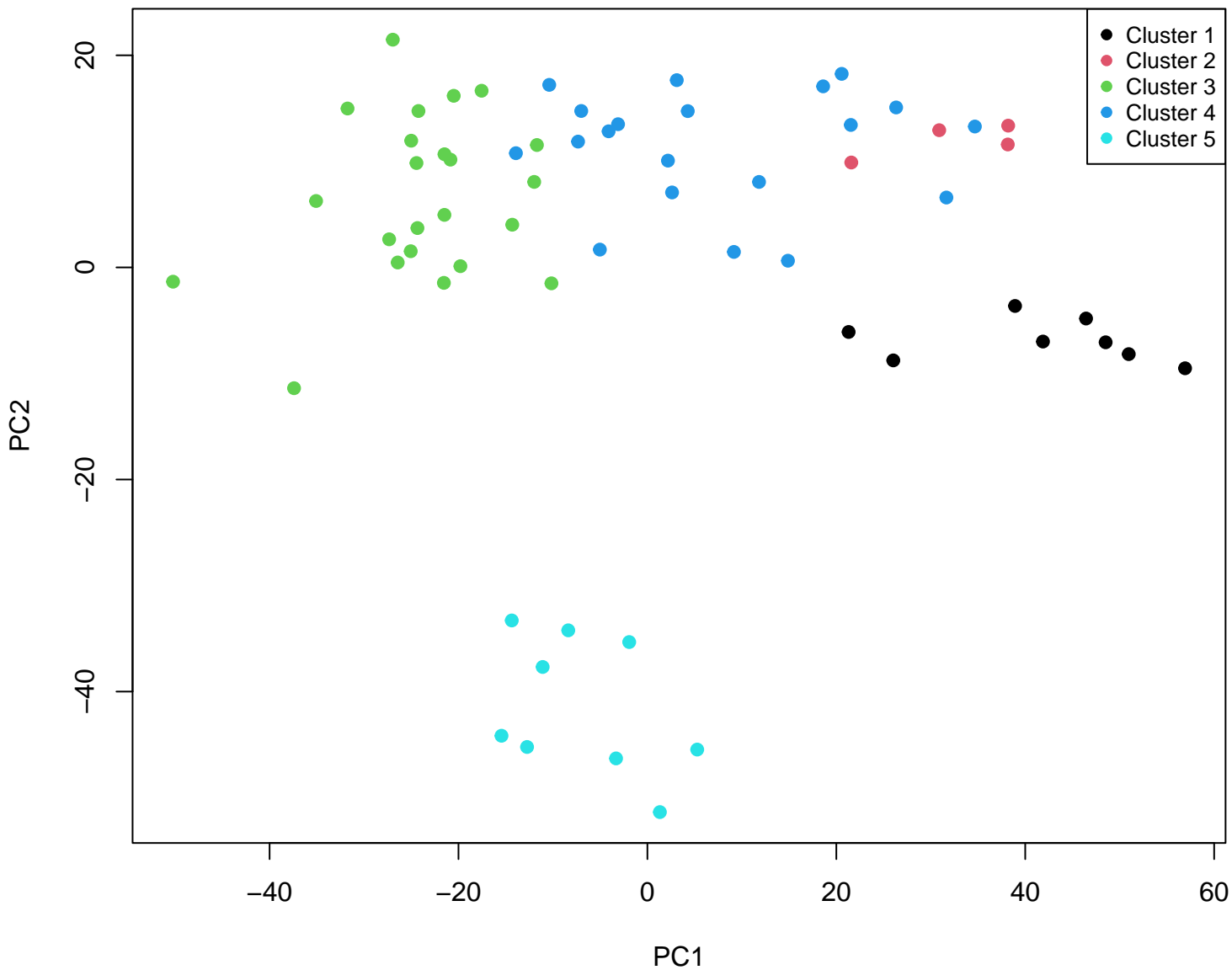
Reduced data Z\_sp k-means clusters with 5 clusters (using 2 randomly sampled genes)



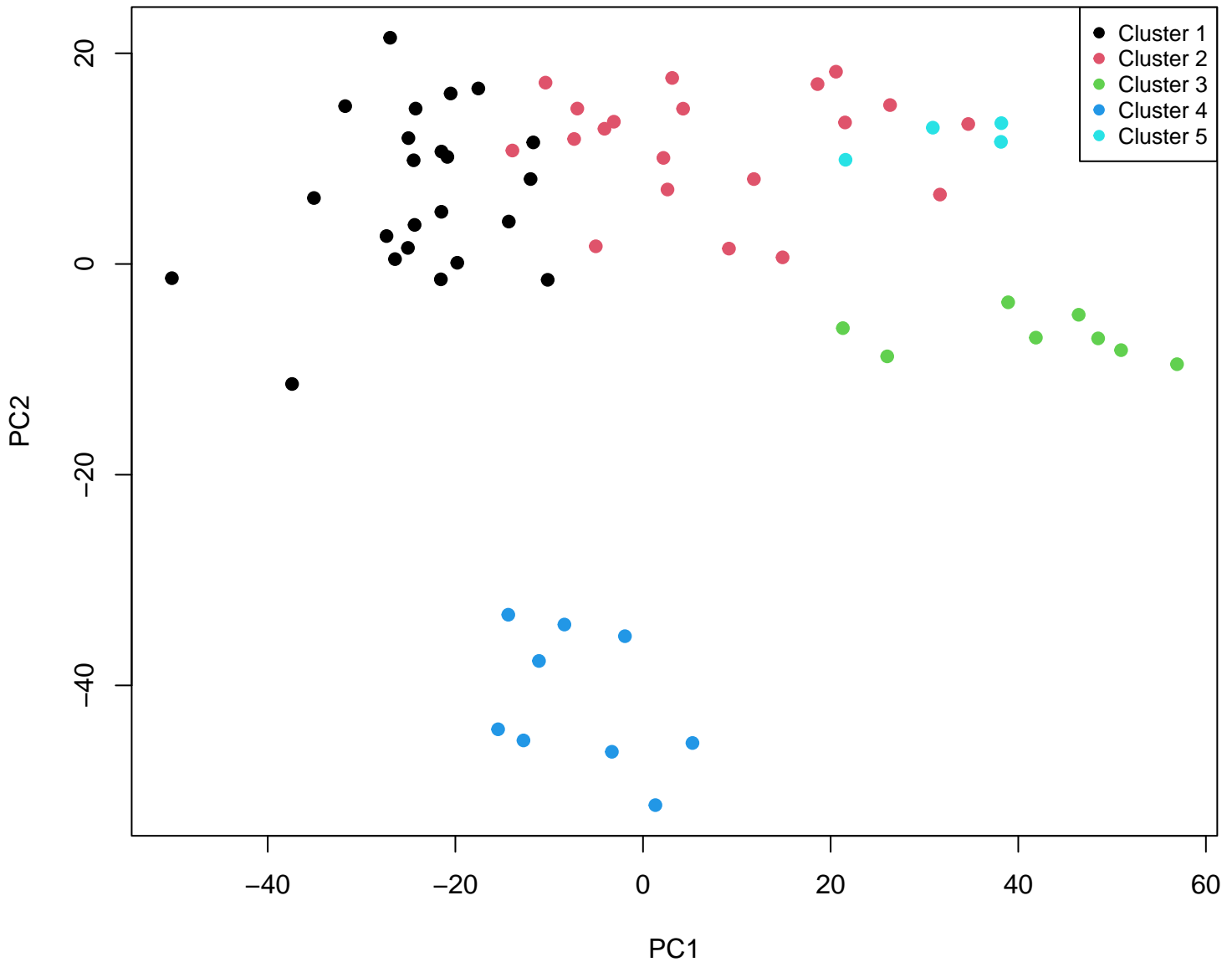
**K-means clustering (k=5) on reduced data Z**



**K-means clustering (k=5) on reduced data Z\_75**

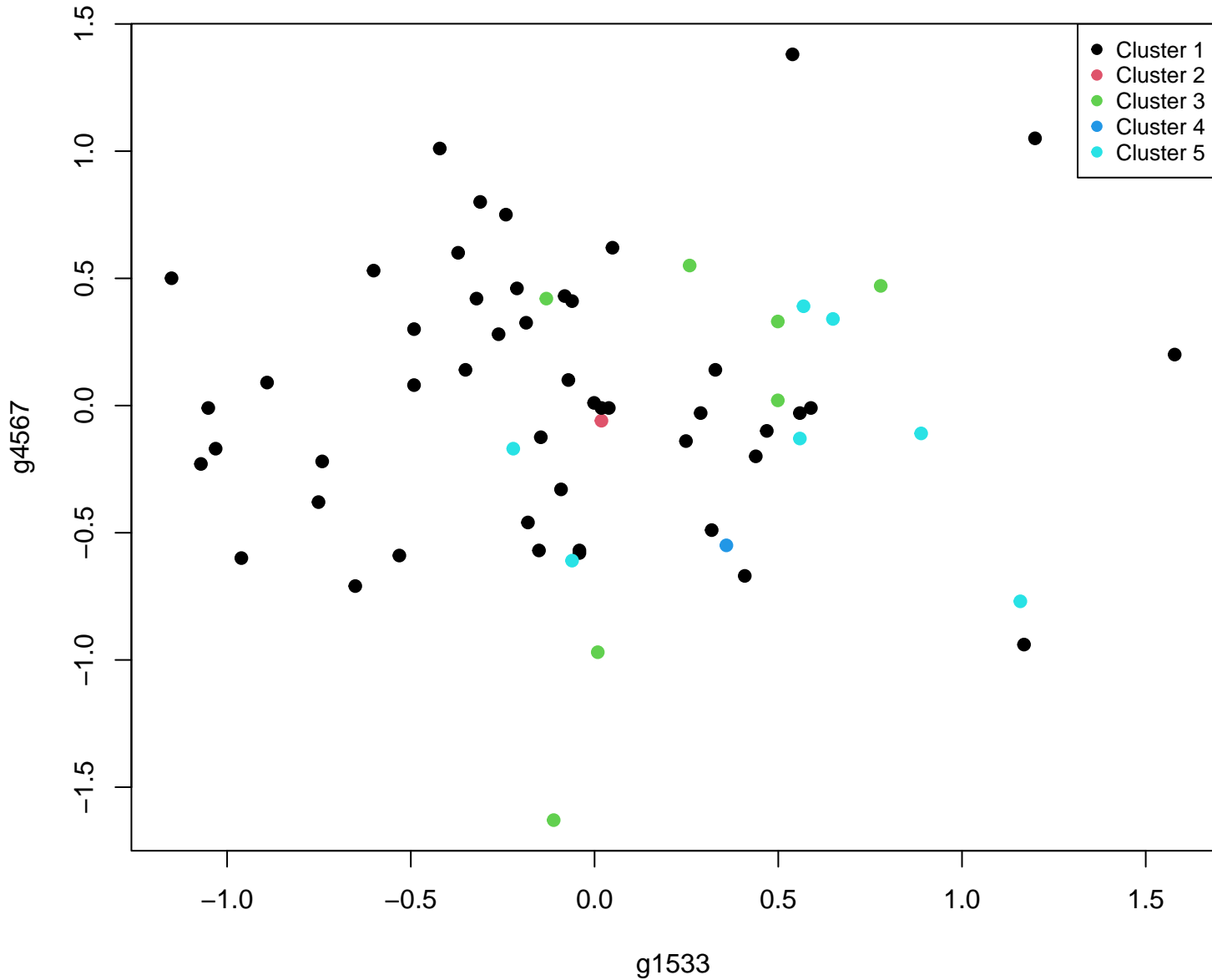


**K-means clustering (k=5) on reduced data Z\_sp**

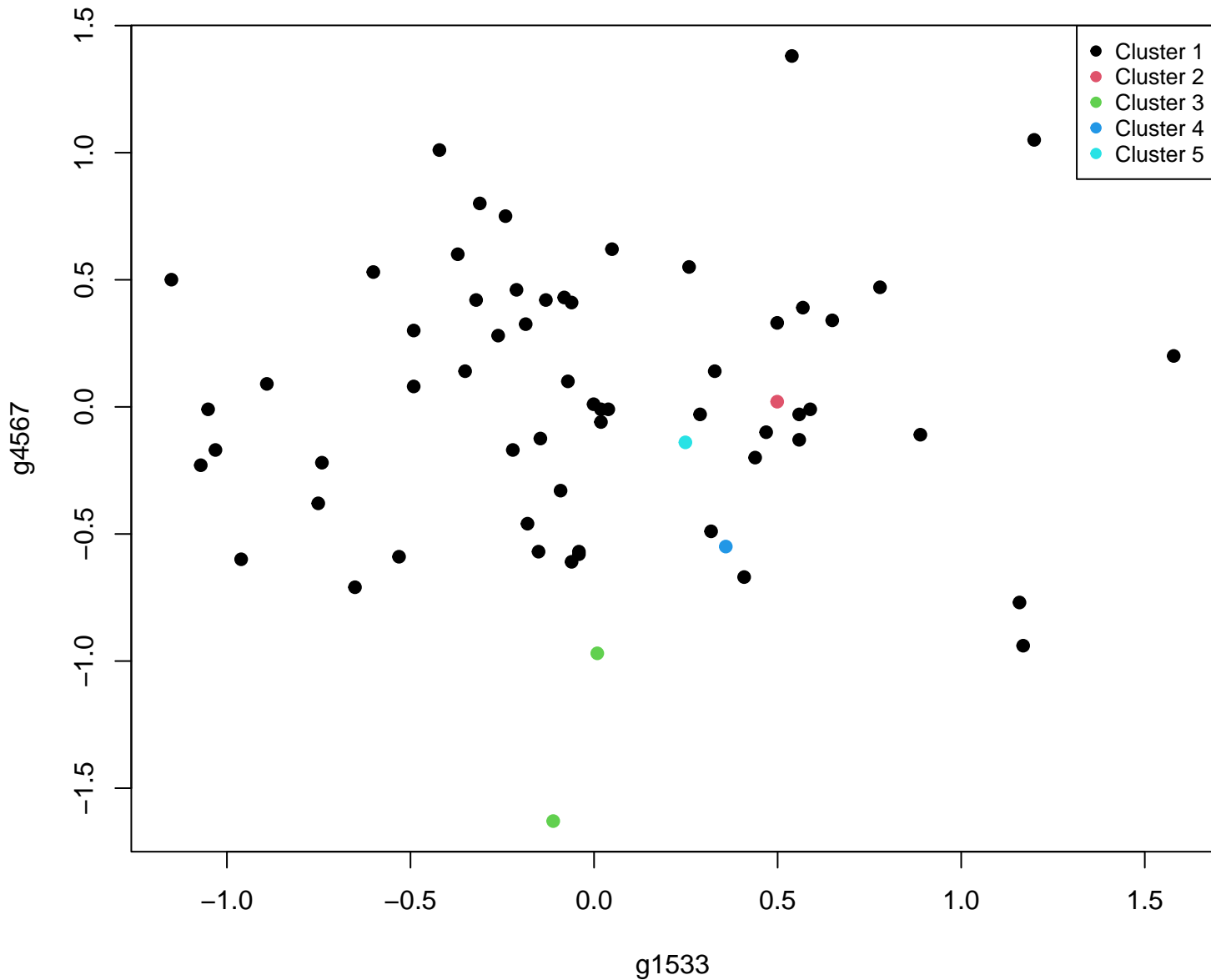




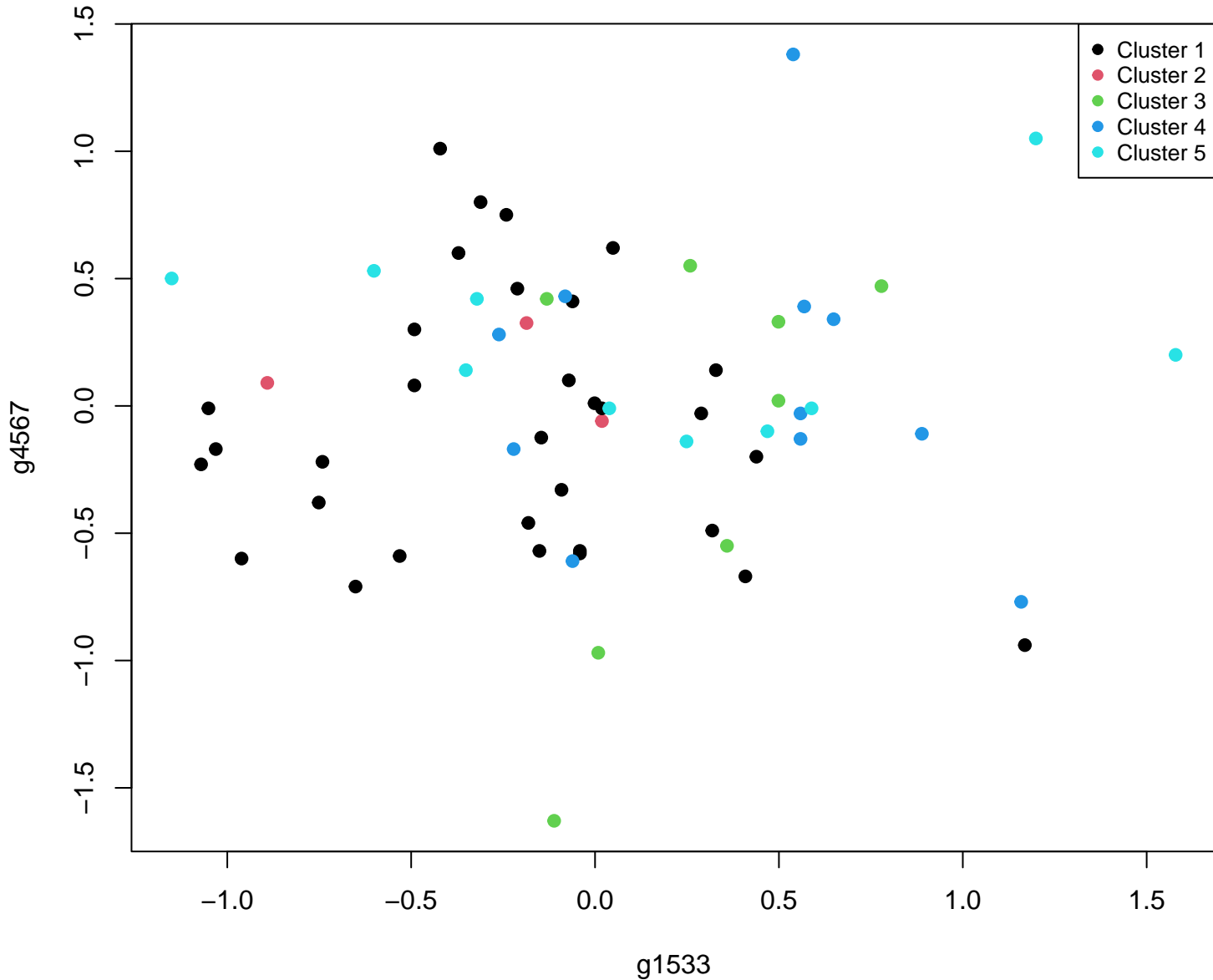
Average linkage with 5 cluster cutoff on Z (sample)



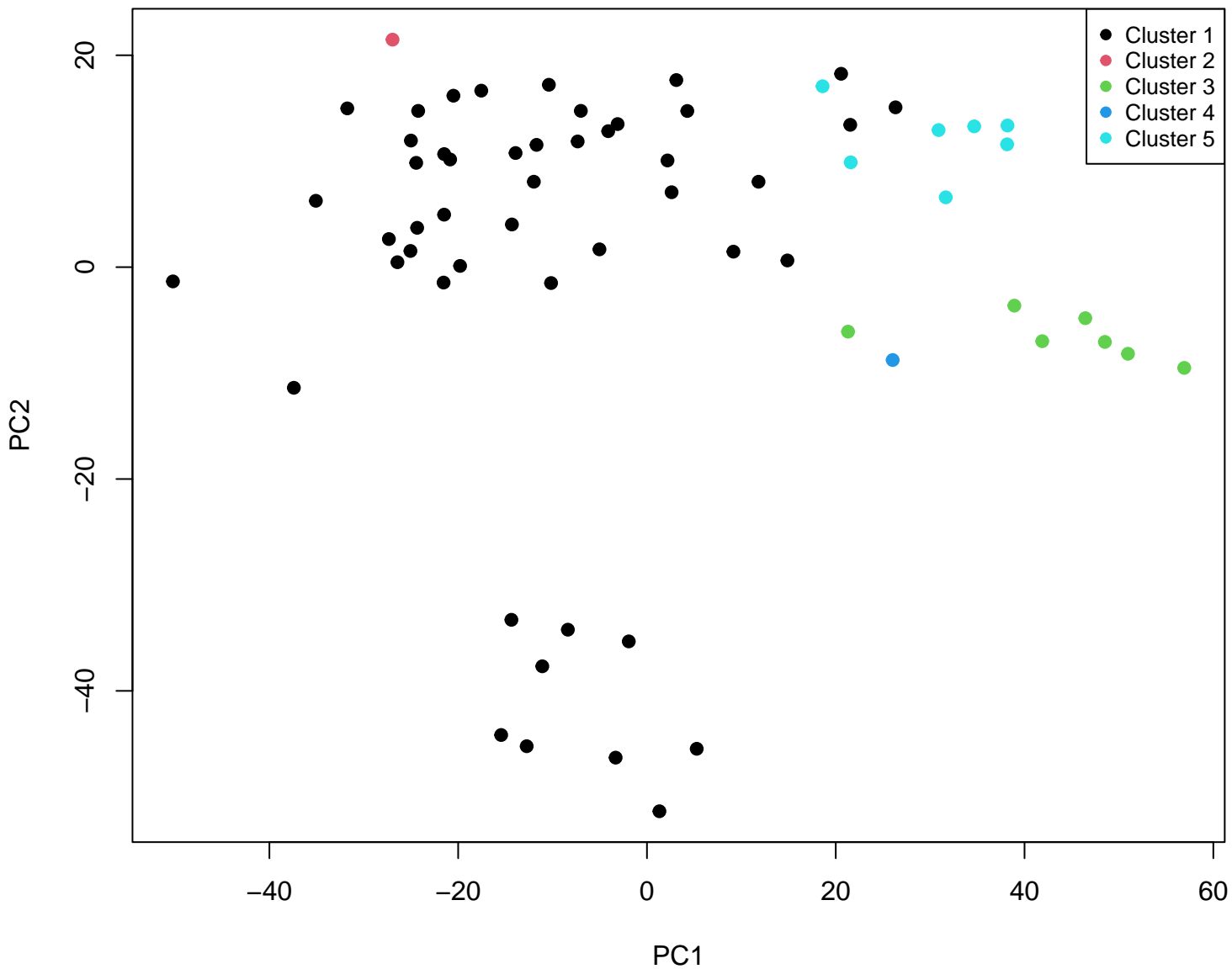
Single linkage with 5 cluster cutoff on Z (sample)



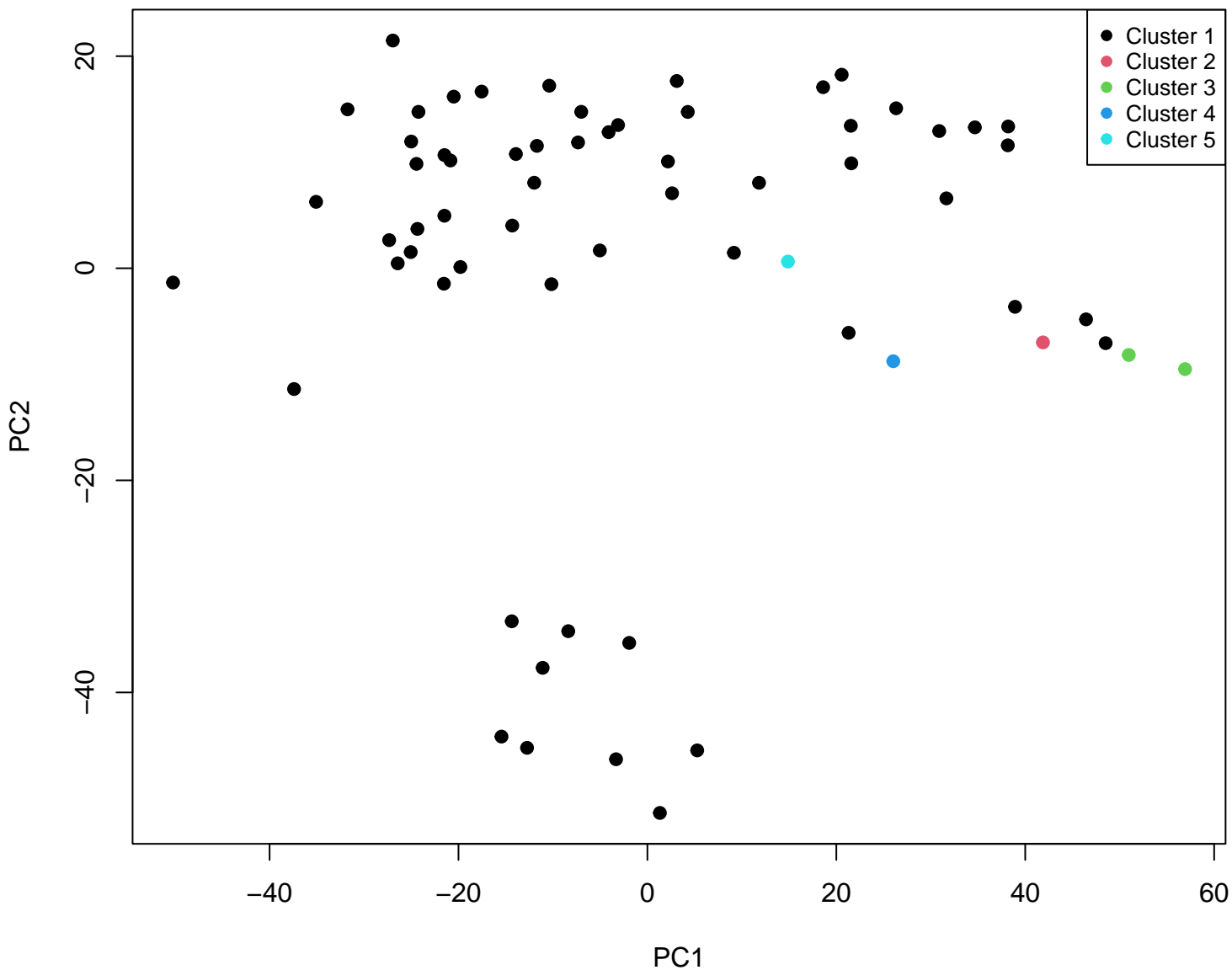
Complete linkage with 5 cluster cutoff on Z (sample)



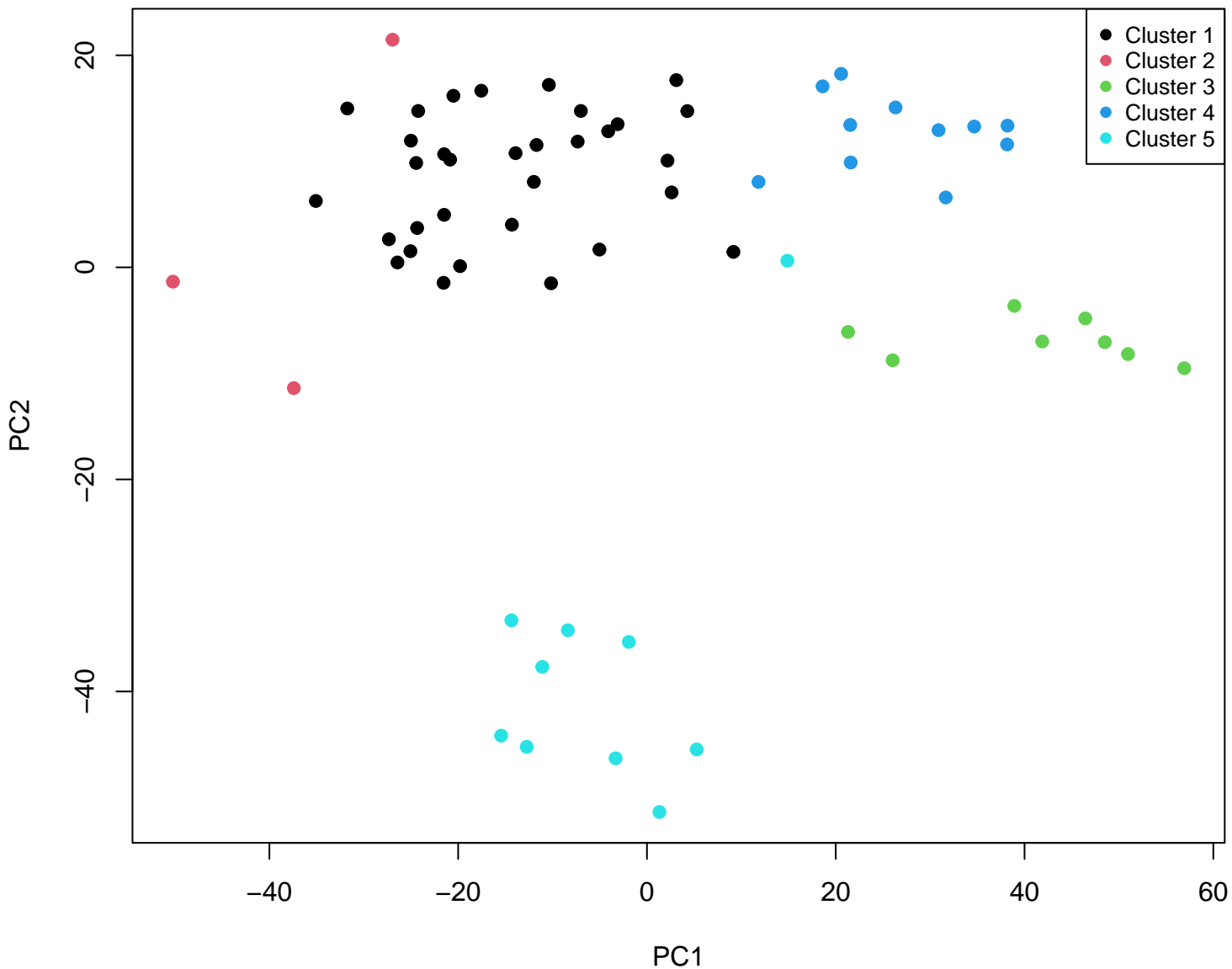
Average linkage clusters (first 2 PCs, 5 cluster cut-off) on Z



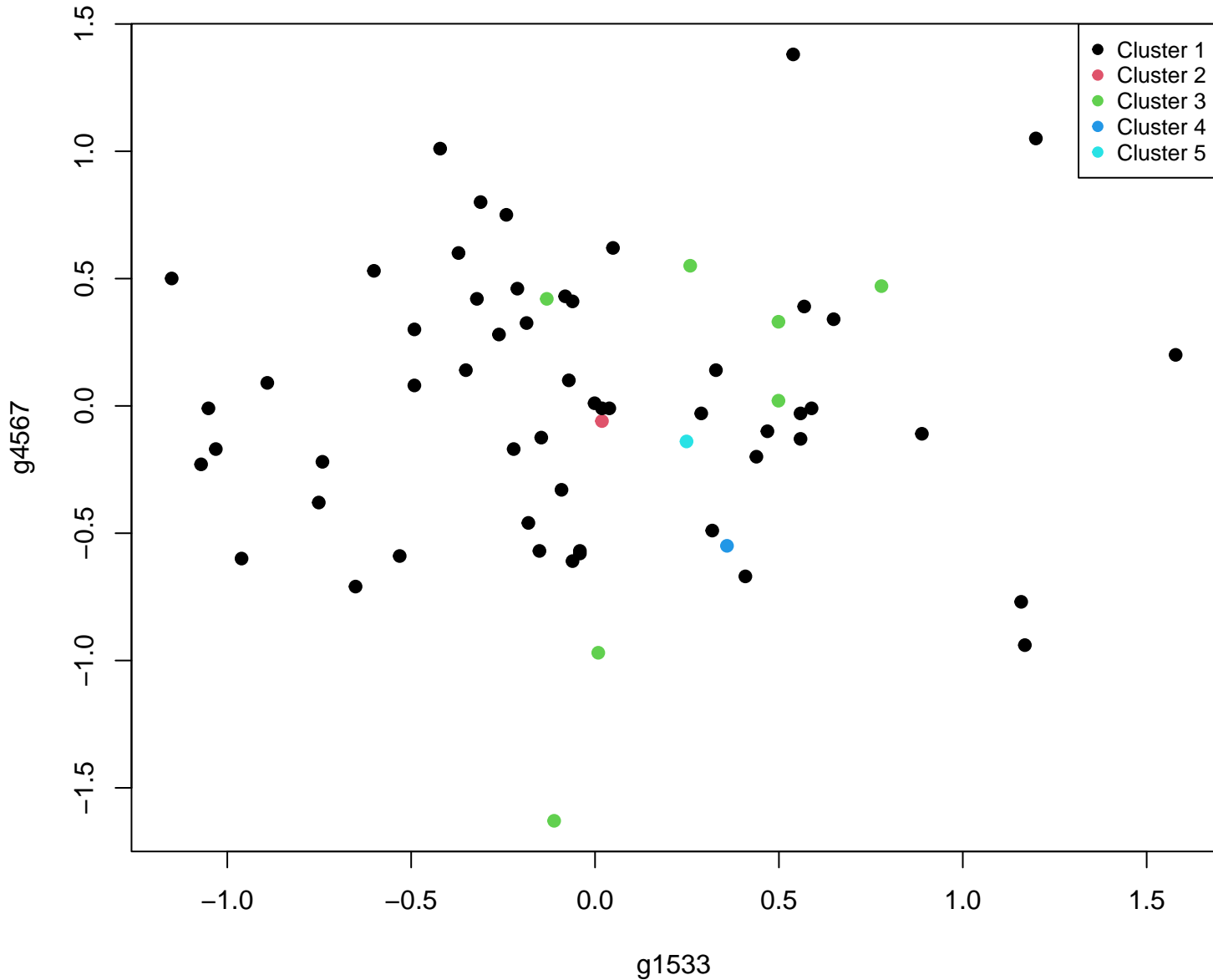
Single linkage clusters (first 2 PCs, 5 cluster cut-off) on Z



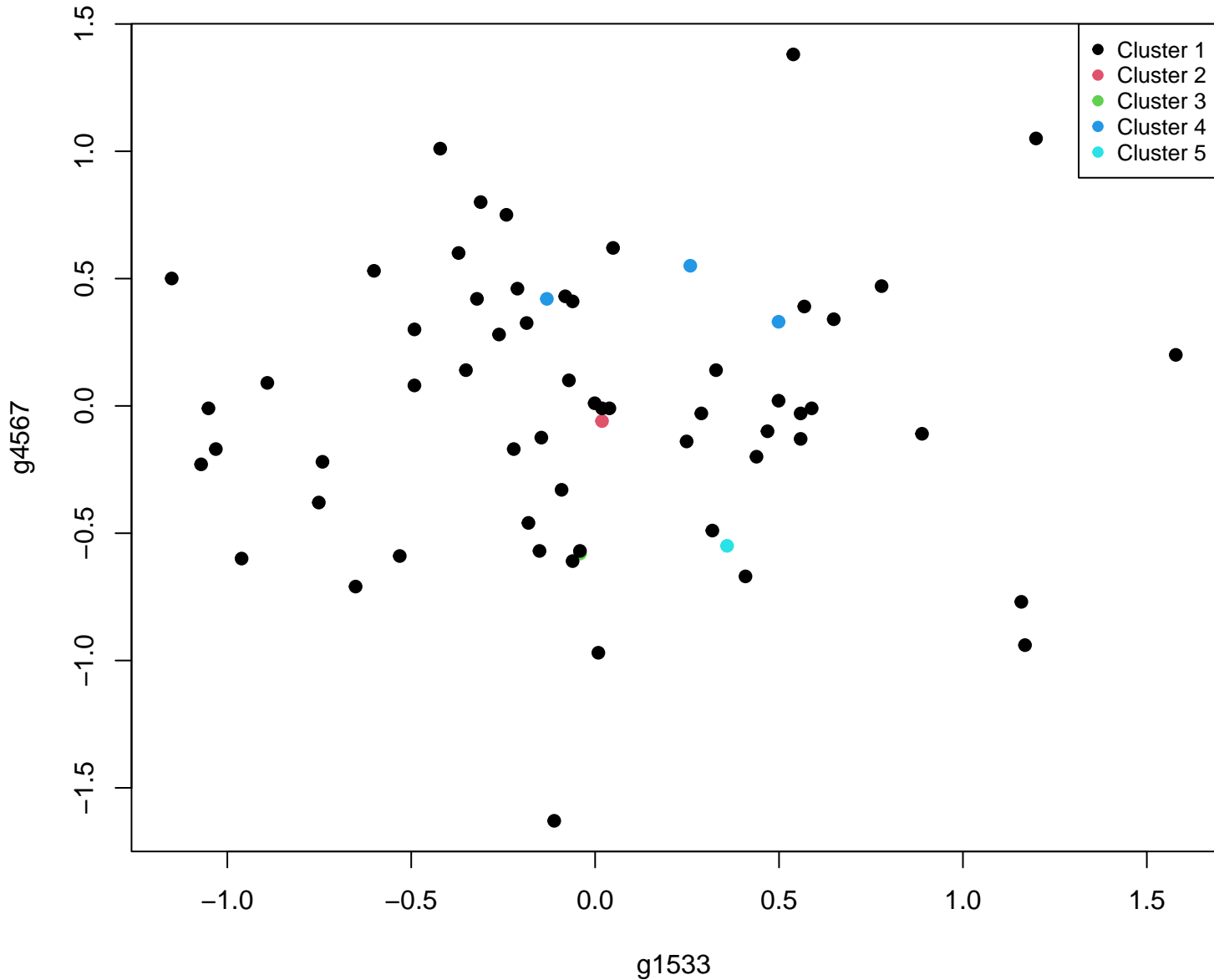
**Complete linkage clusters (first 2 PCs, 5 cluster cut-off) on Z**



Average linkage with 5 cluster cutoff on Z\_75

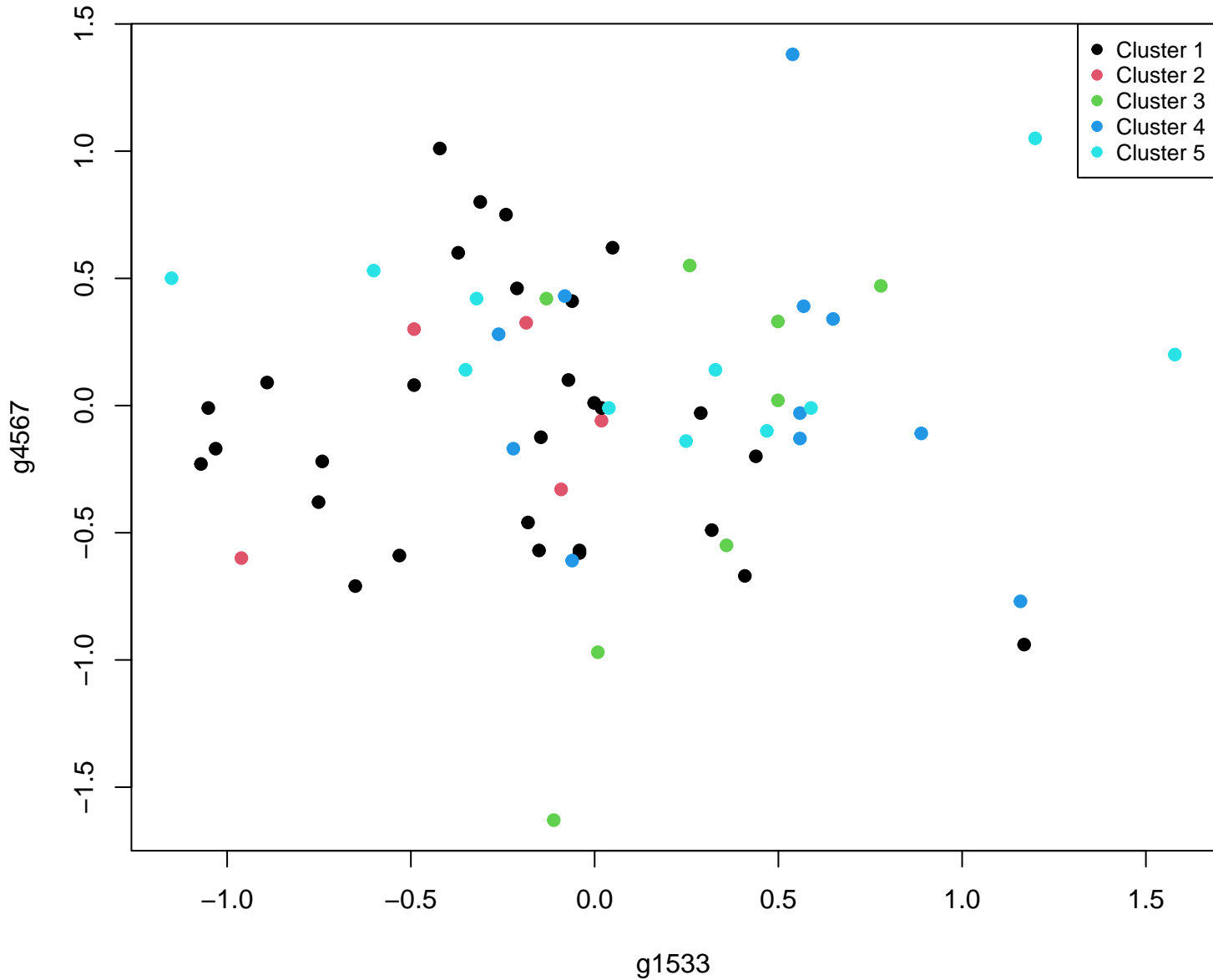


Single linkage with 5 cluster cutoff on Z\_75

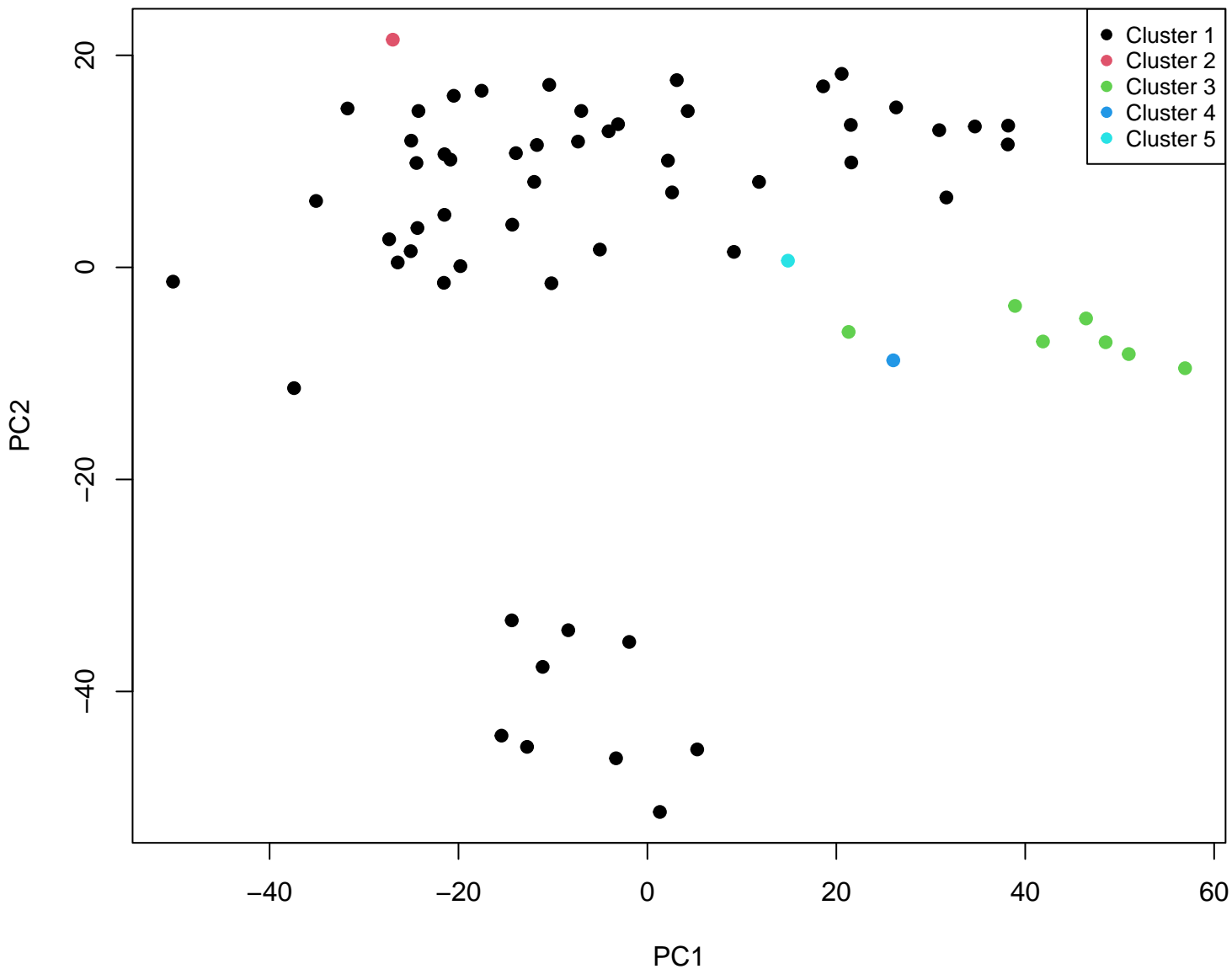




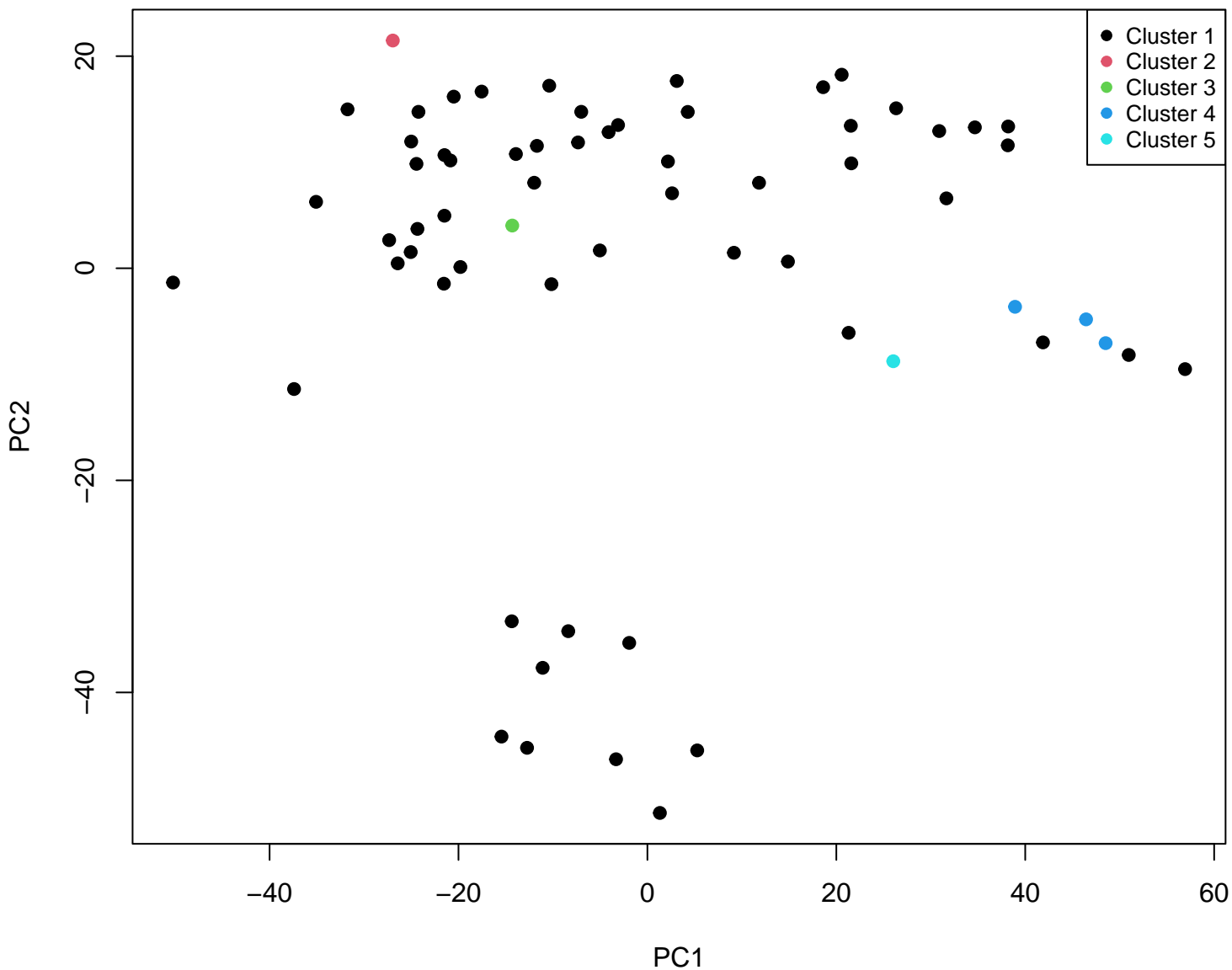
Complete linkage with 5 cluster cutoff on Z\_75



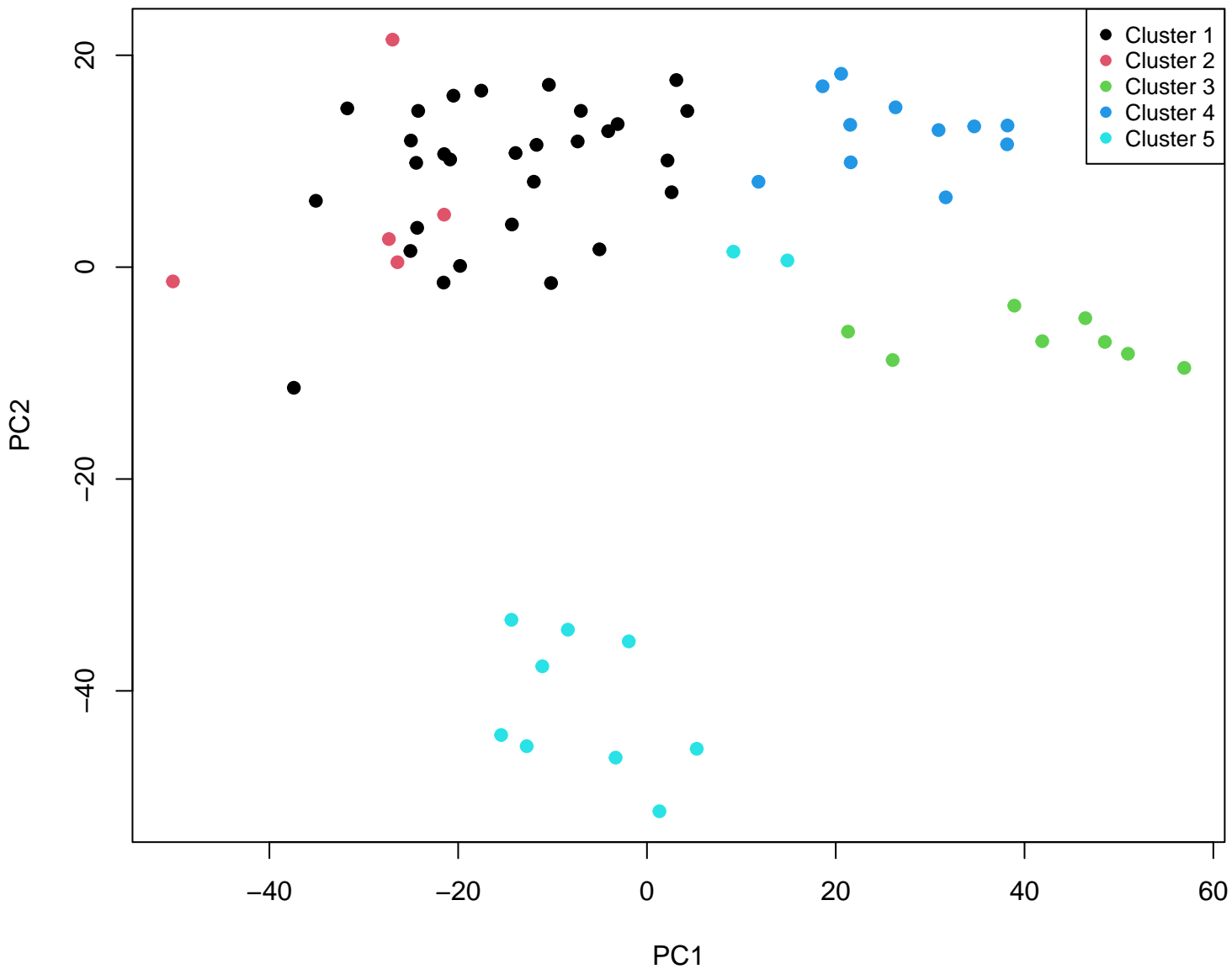
Average linkage clusters (first 2 PCs, 5 cluster cut-off) on Z\_75



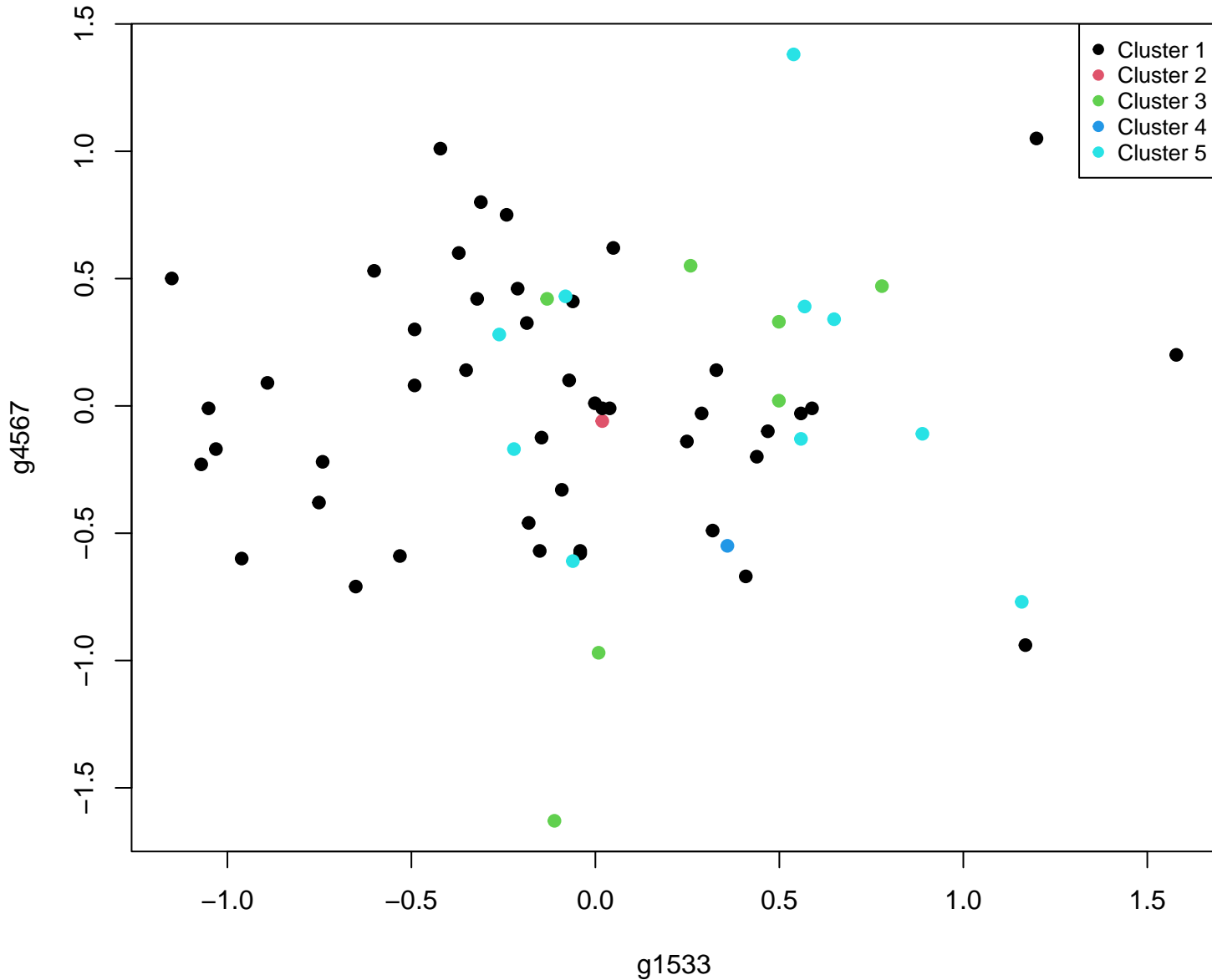
Single linkage clusters (first 2 PCs, 5 cluster cut-off) on Z\_75



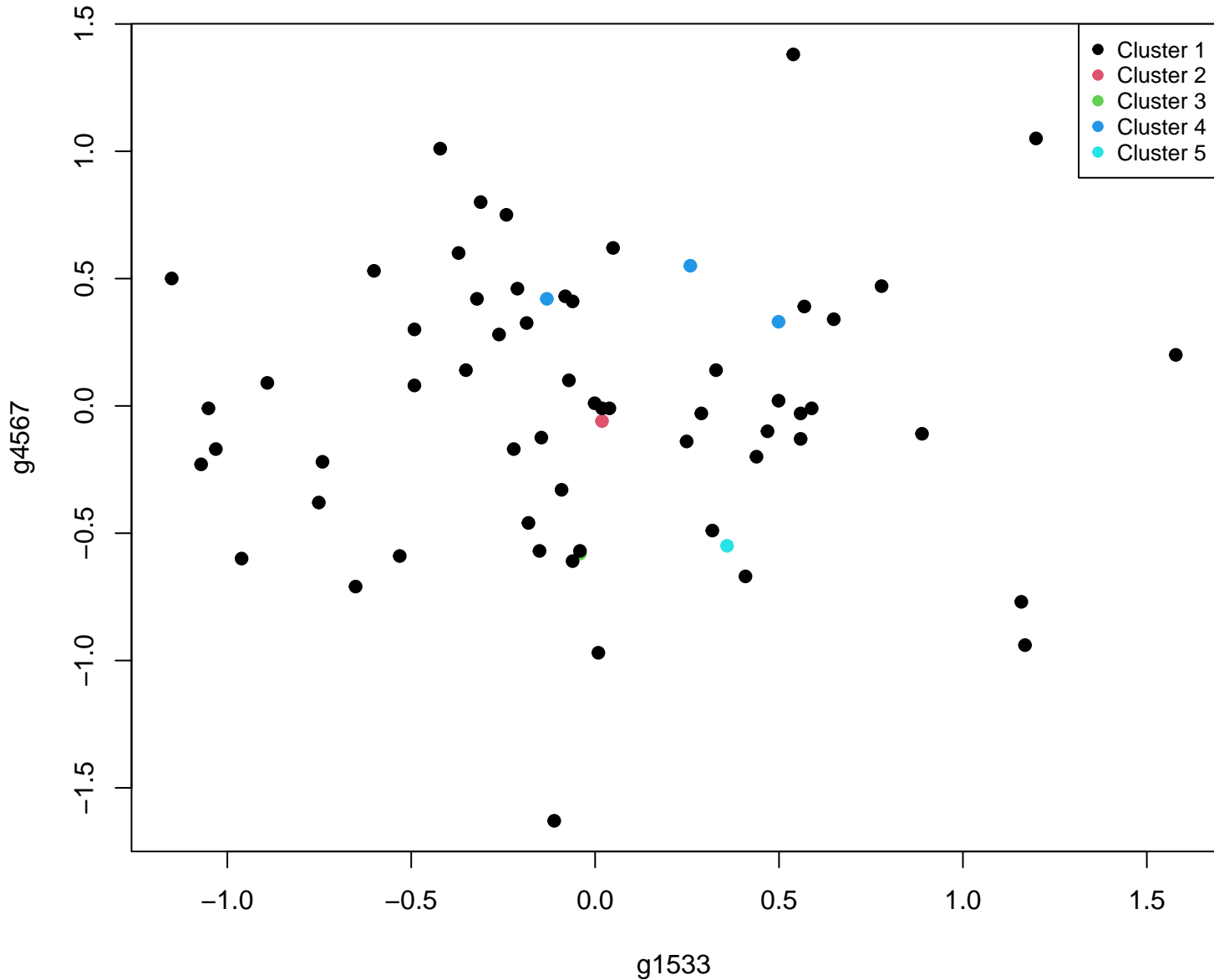
**Complete linkage clusters (first 2 PCs, 5 cluster cut-off) on Z\_75**



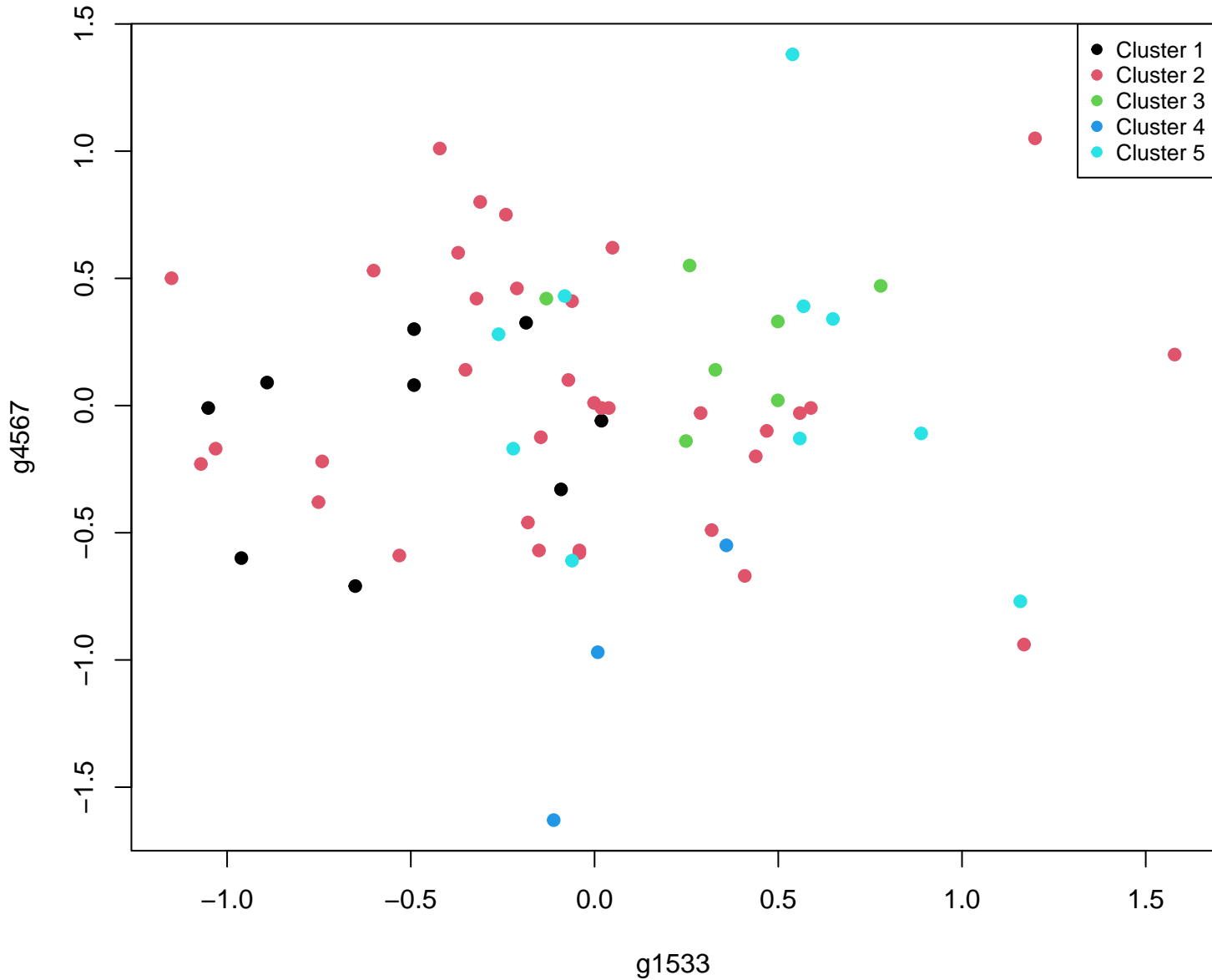
Average linkage with 5 cluster cutoff on Z\_sp



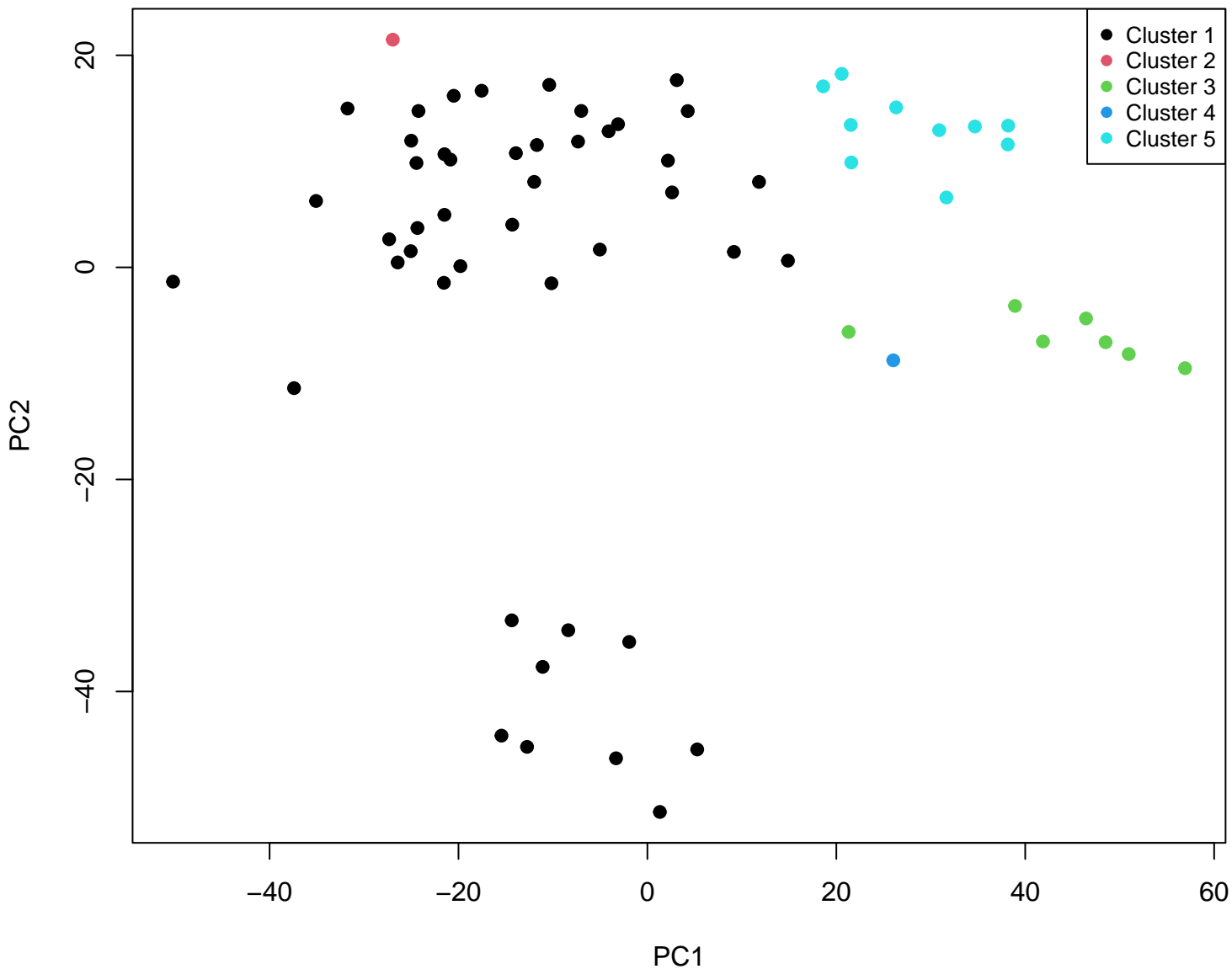
Single linkage with 5 cluster cutoff on Z\_sp



Complete linkage with 5 cluster cutoff on Z\_sp

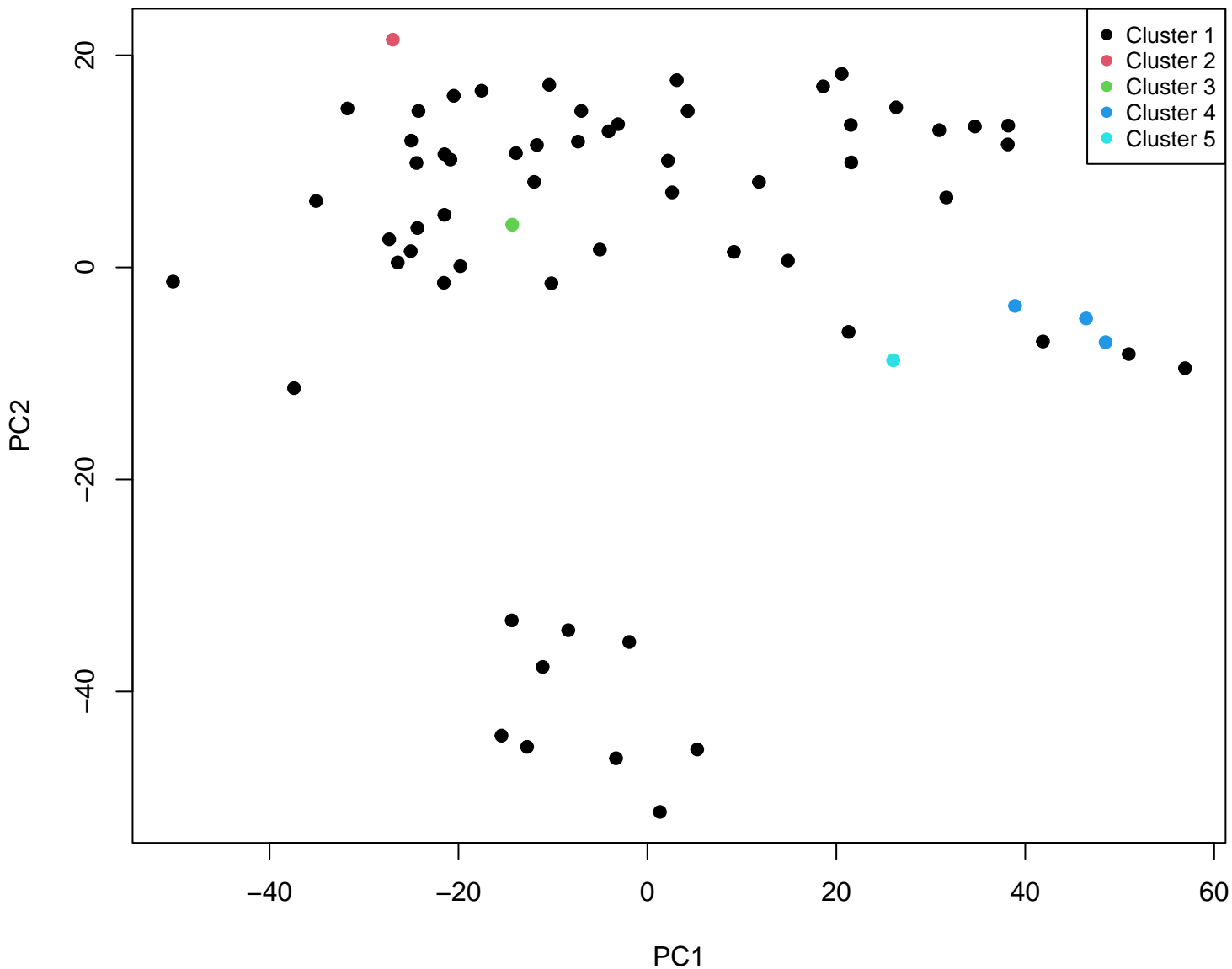


Average linkage clusters (first 2 PCs, 5 cluster cut-off) on Z\_sp

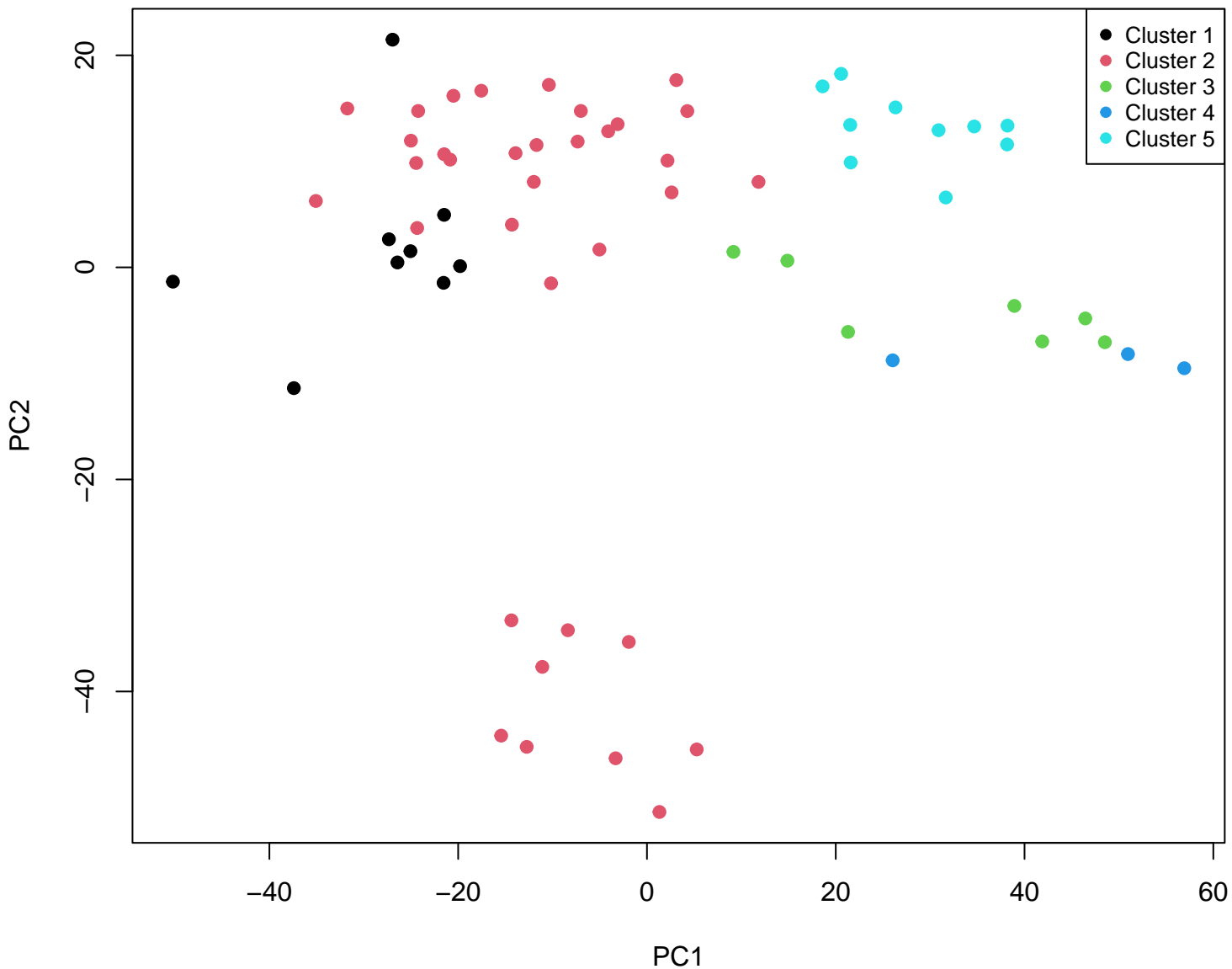




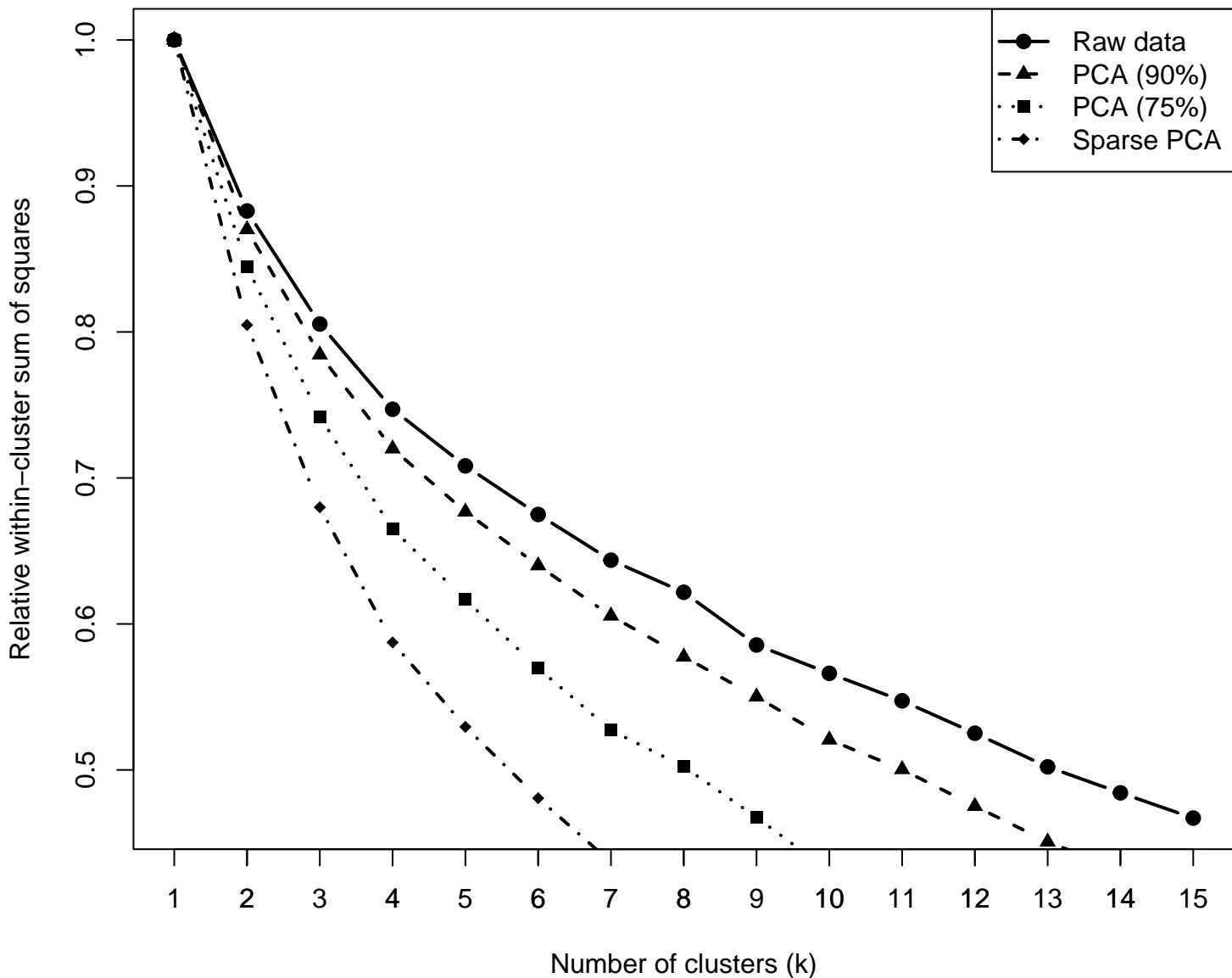
### Single linkage clusters (first 2 PCs, 5 cluster cut-off) on Z\_sp



Complete linkage clusters (first 2 PCs, 5 cluster cut-off) on Z\_sp



Normalised elbow plots for k-means



ARI vs number of clusters for k-means

