

# RNA-Seq Sample QC

**Before sample removal**

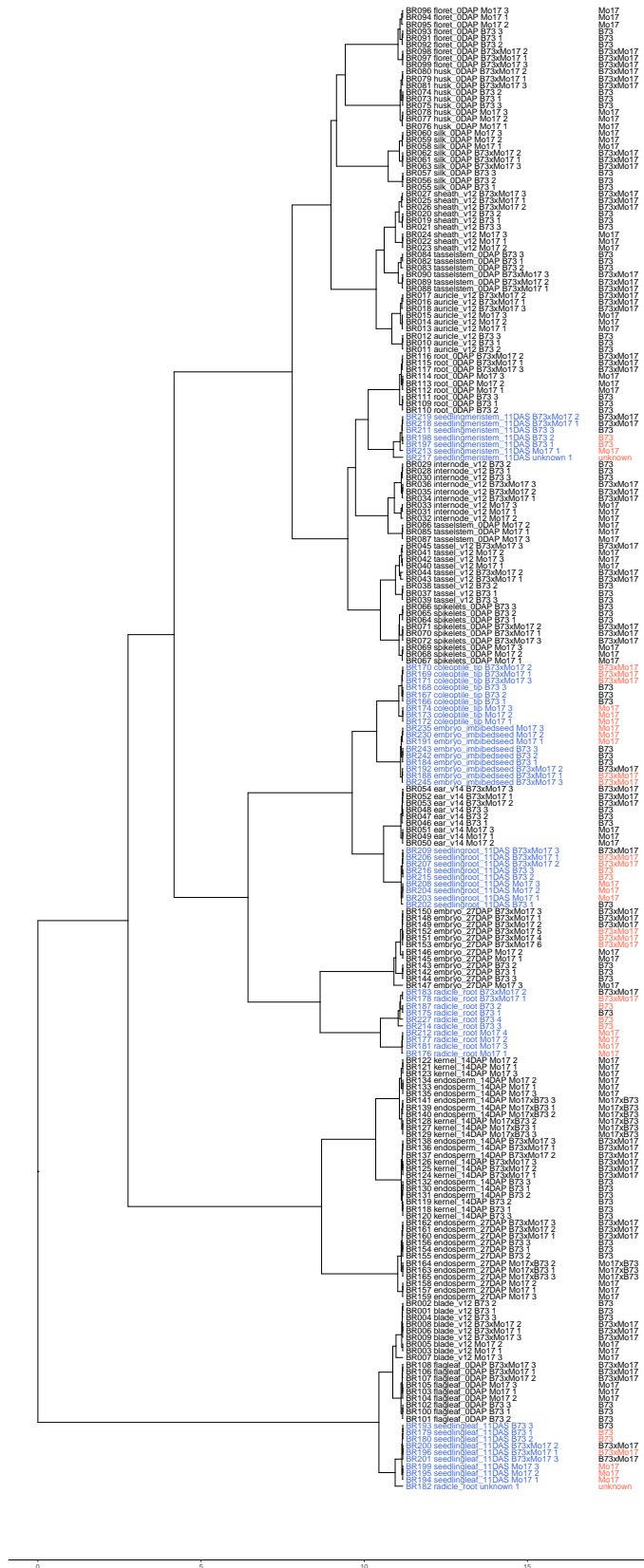
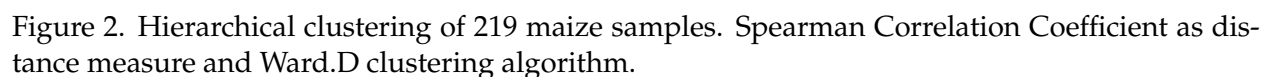


Figure 1. Hierarchical clustering of 219 maize samples. Pearson Correlation Coefficient as distance measure and Ward.D clustering algorithm.



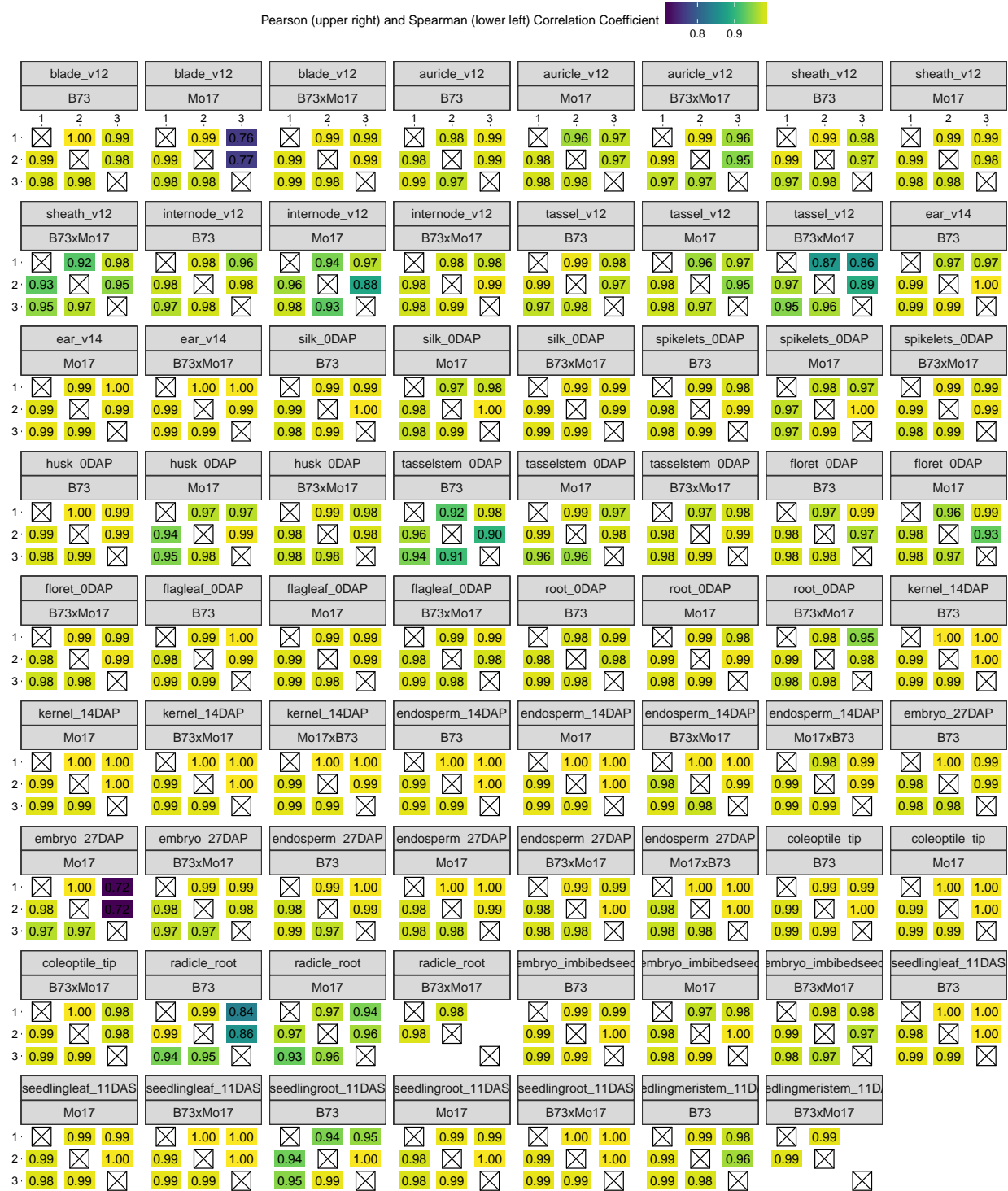


Figure 3. Correlation of FPKM values between each two replicates for each tissue-genotype combination. Only 14,746 genes with FPKM >= 1 in at least 80% samples were used in this analysis.

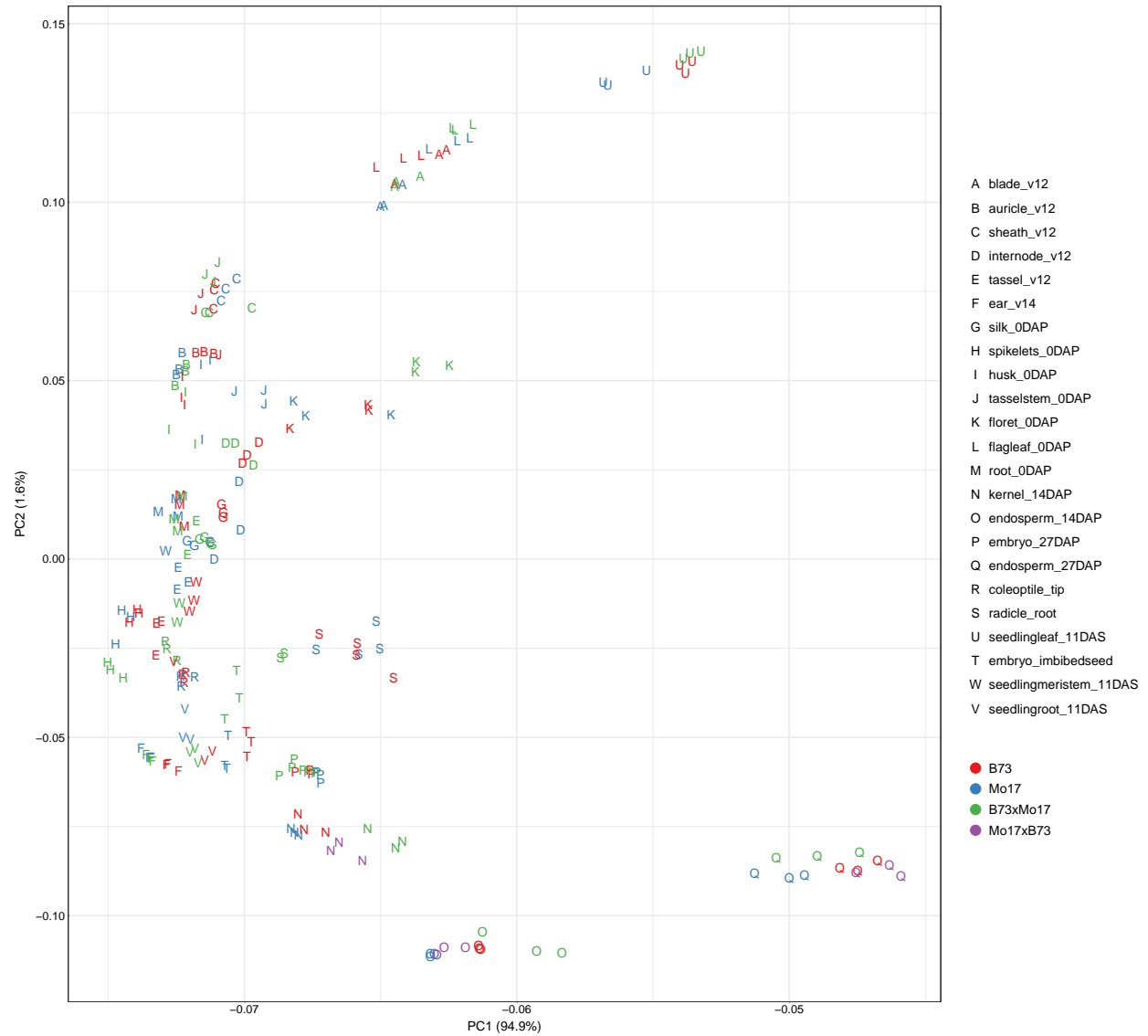
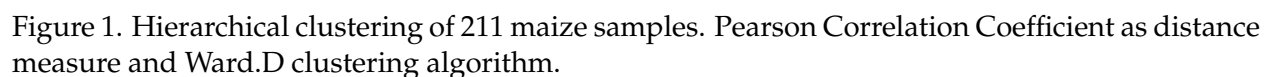


Figure 4. Principle component plot of the maize RNA-Seq expression matrix. The original 39,005 genes were filtered to 14,746 genes that with FPKM  $\geq 1$  in at least 176 out of 219 samples.

### **After sample removal**

- Too few reads: BR182, BR217
- PCC with other replicates too low: BR007, BR045, BR147, BR227, BR214, BR202



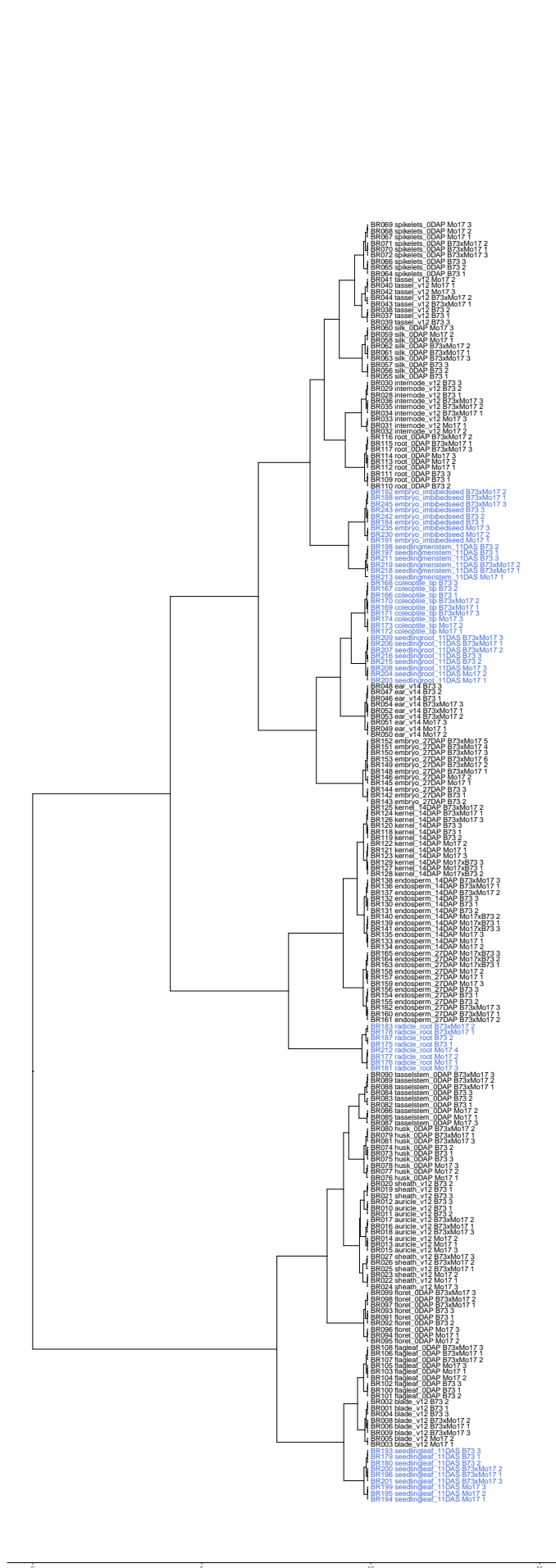


Figure 2. Hierarchical clustering of 211 maize samples. Spearman Correlation Coefficient as distance measure and Ward.D clustering algorithm.



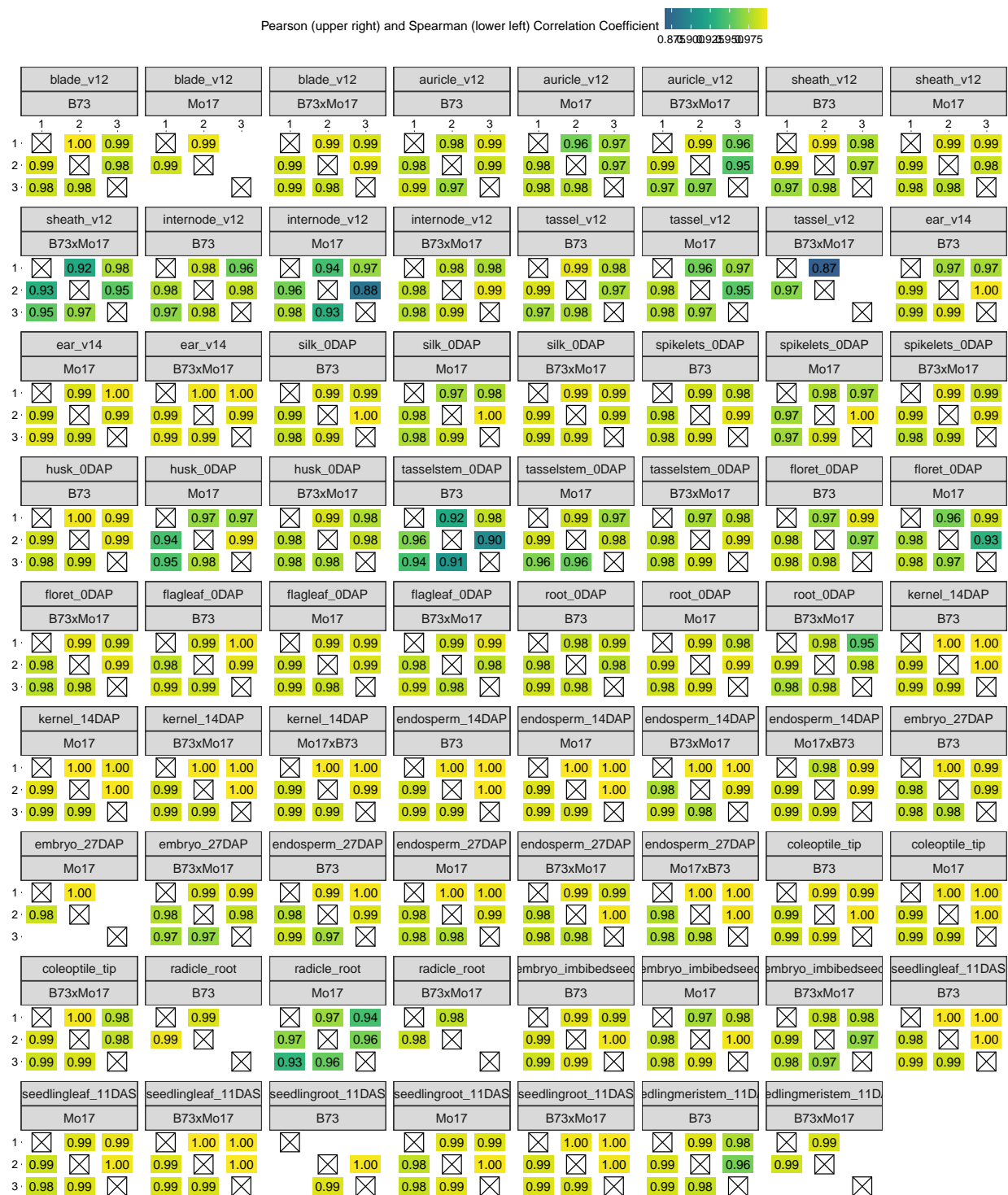


Figure 3. Correlation of FPKM values between each two replicates for each tissue-genotype combination. Only 14,746 genes with FPKM  $\geq 1$  in at least 170 (80%) out of 211 samples were used in this analysis.

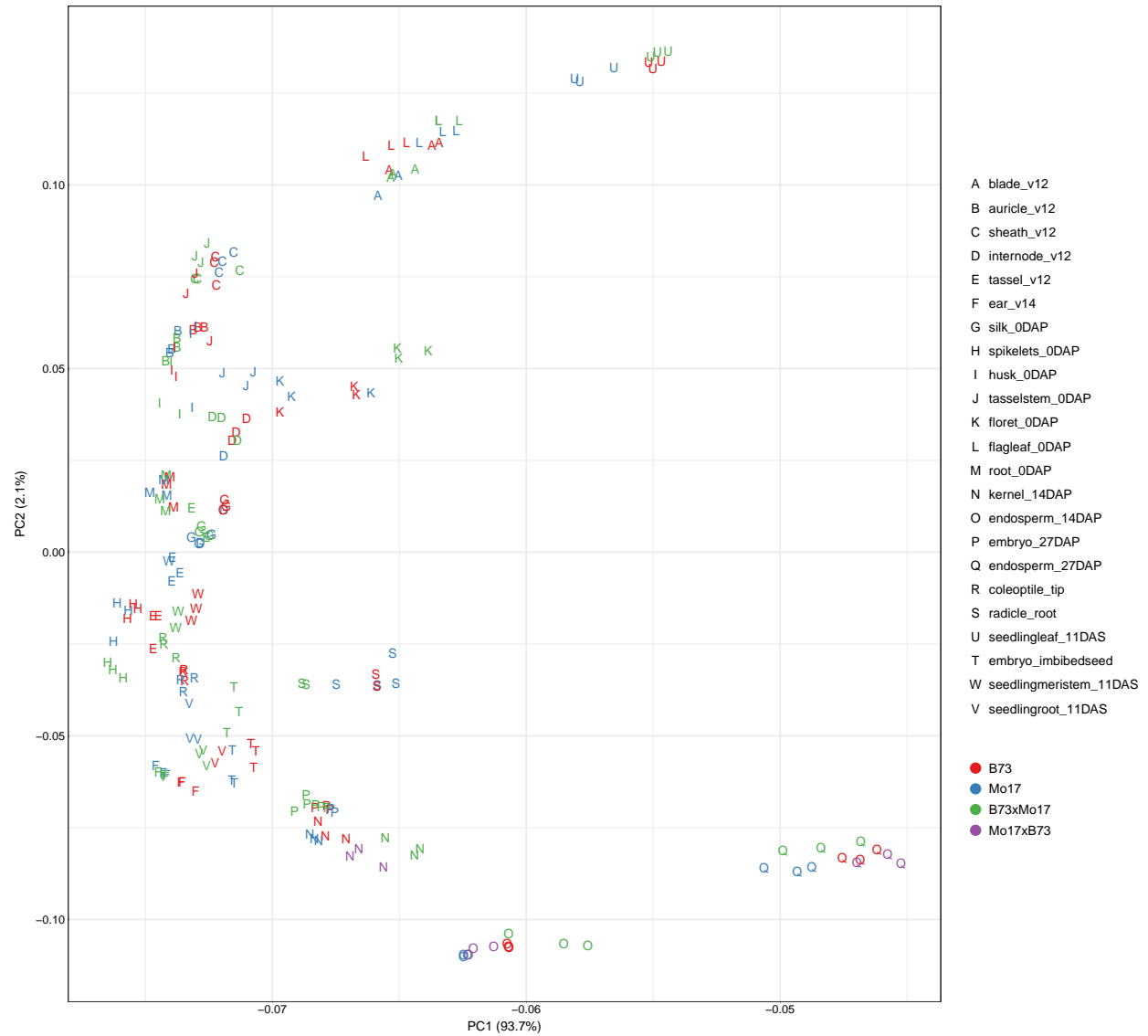


Figure 4. Principle component plot of the maize RNA-Seq expression matrix. The original 39,005 genes were filtered to 14,746 genes that with FPKM  $\geq 1$  in at least 170 (80%) out of 211 samples.