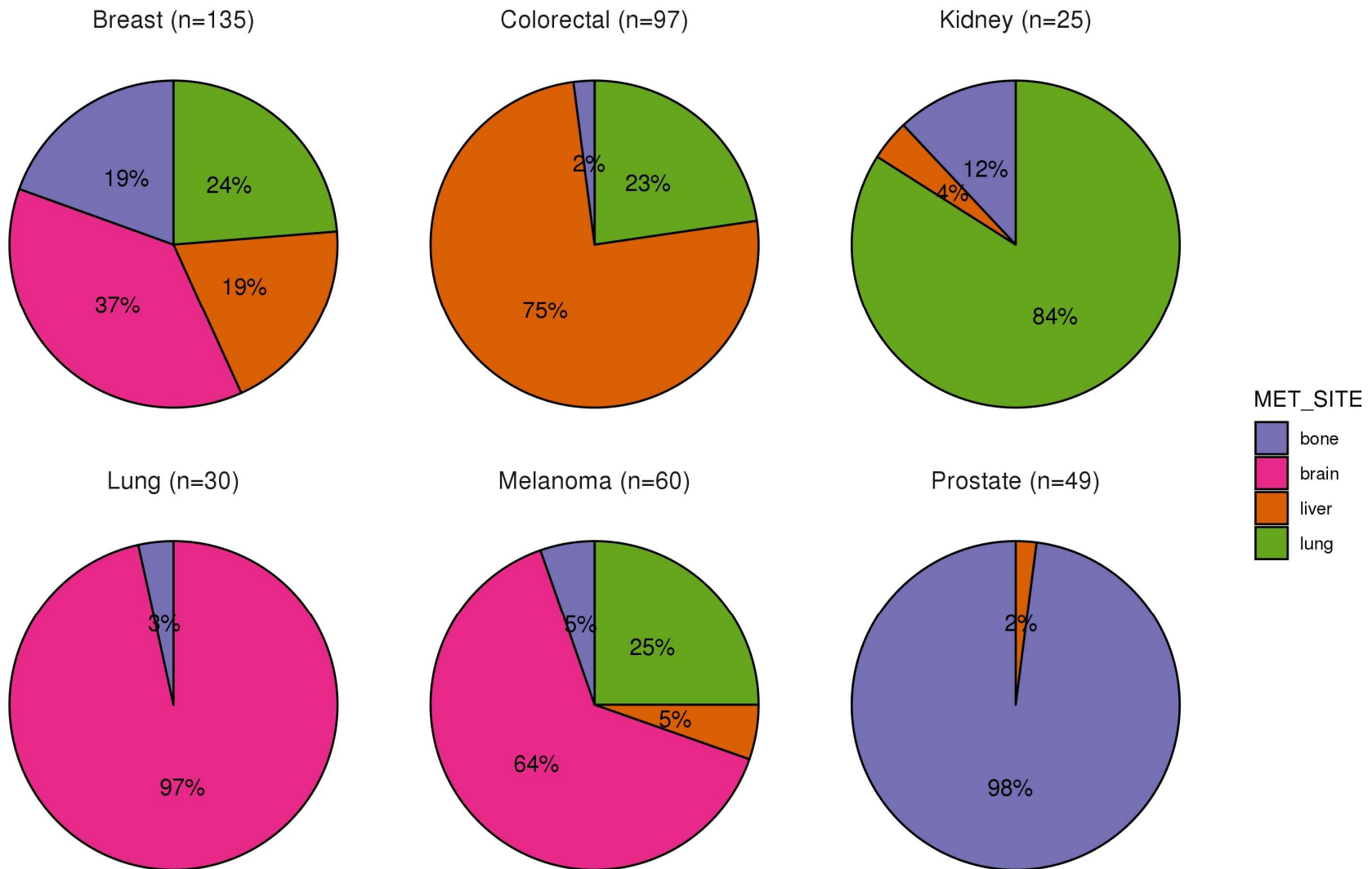
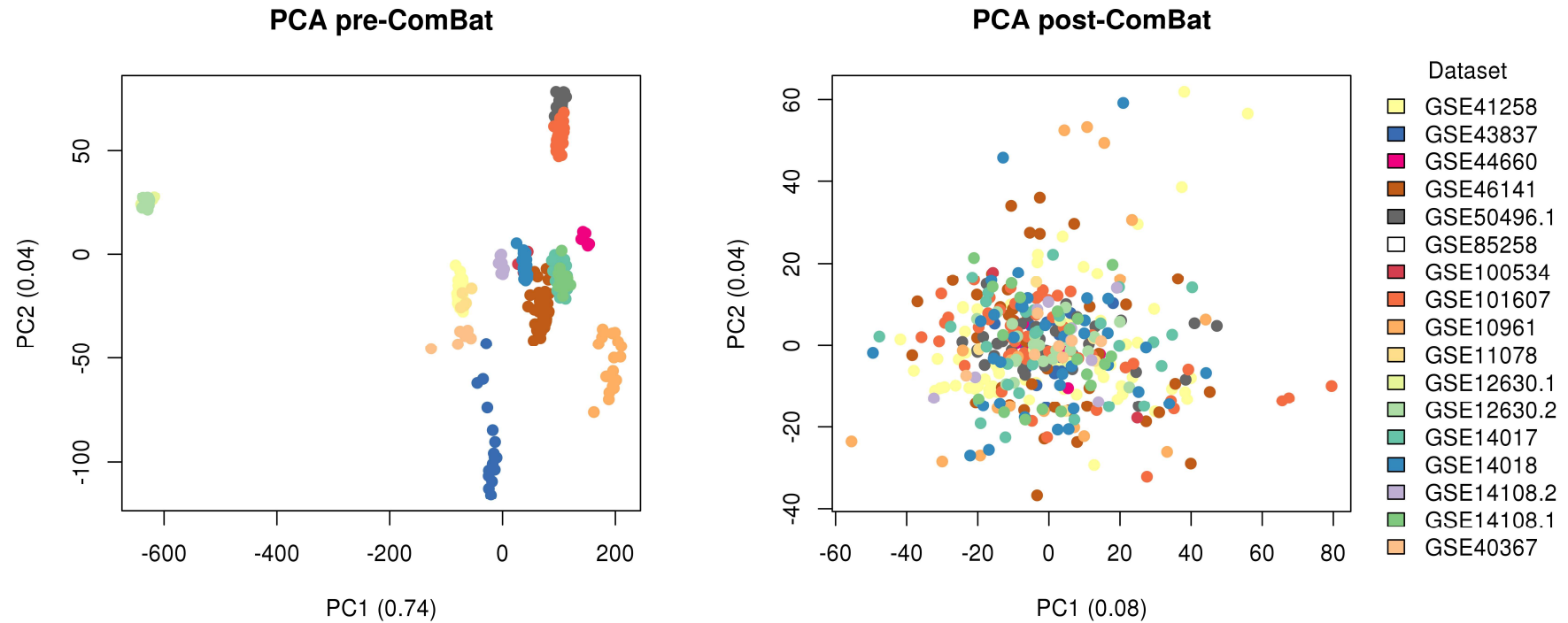


## Supplementary Figure 1



**Supplementary Figure 1.** Piechart showing percentages of metastatic location distribution through the six primary cancer types used in the study. NSCLC: non-small cell lung cancer.

## Supplementary Figure 2



**Supplementary Figure 2.** Principal components analysis (PCA) comparing pre and post-adjustment using ComBat.

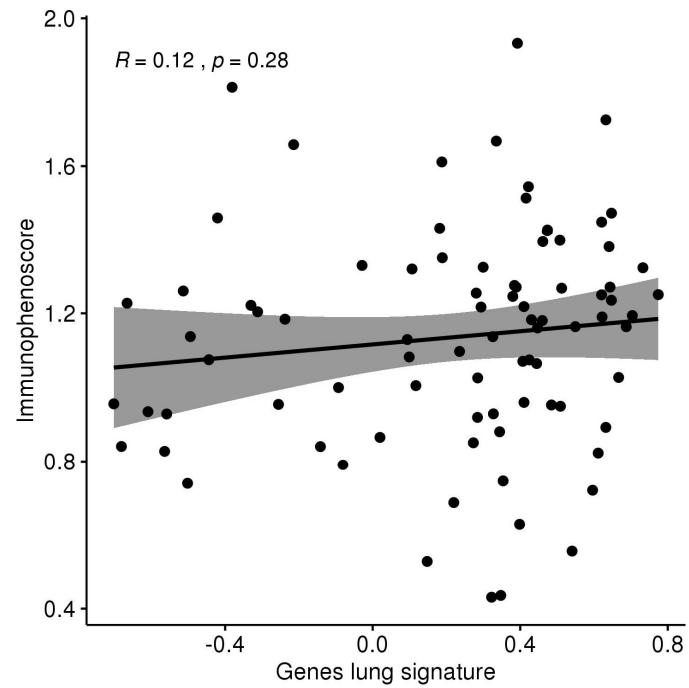
## Supplementary Figure 3



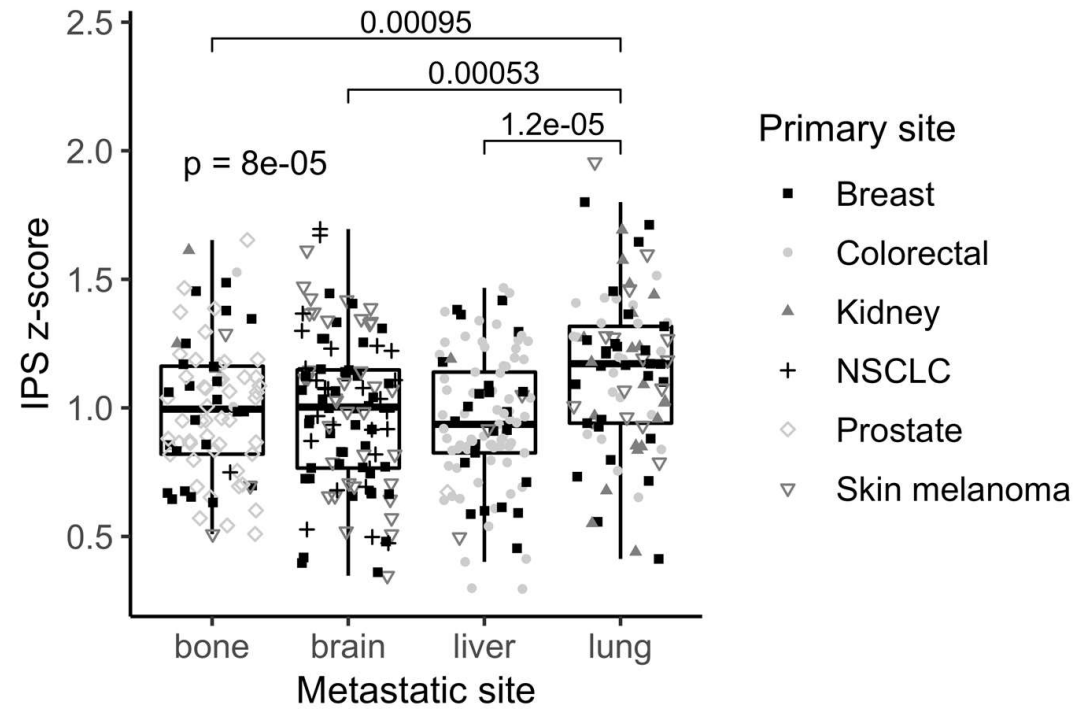
**Supplementary Figure 3.** Heatmap showing healthy tissue gene expression signatures score. For better representation, scores were summarized by sample type.

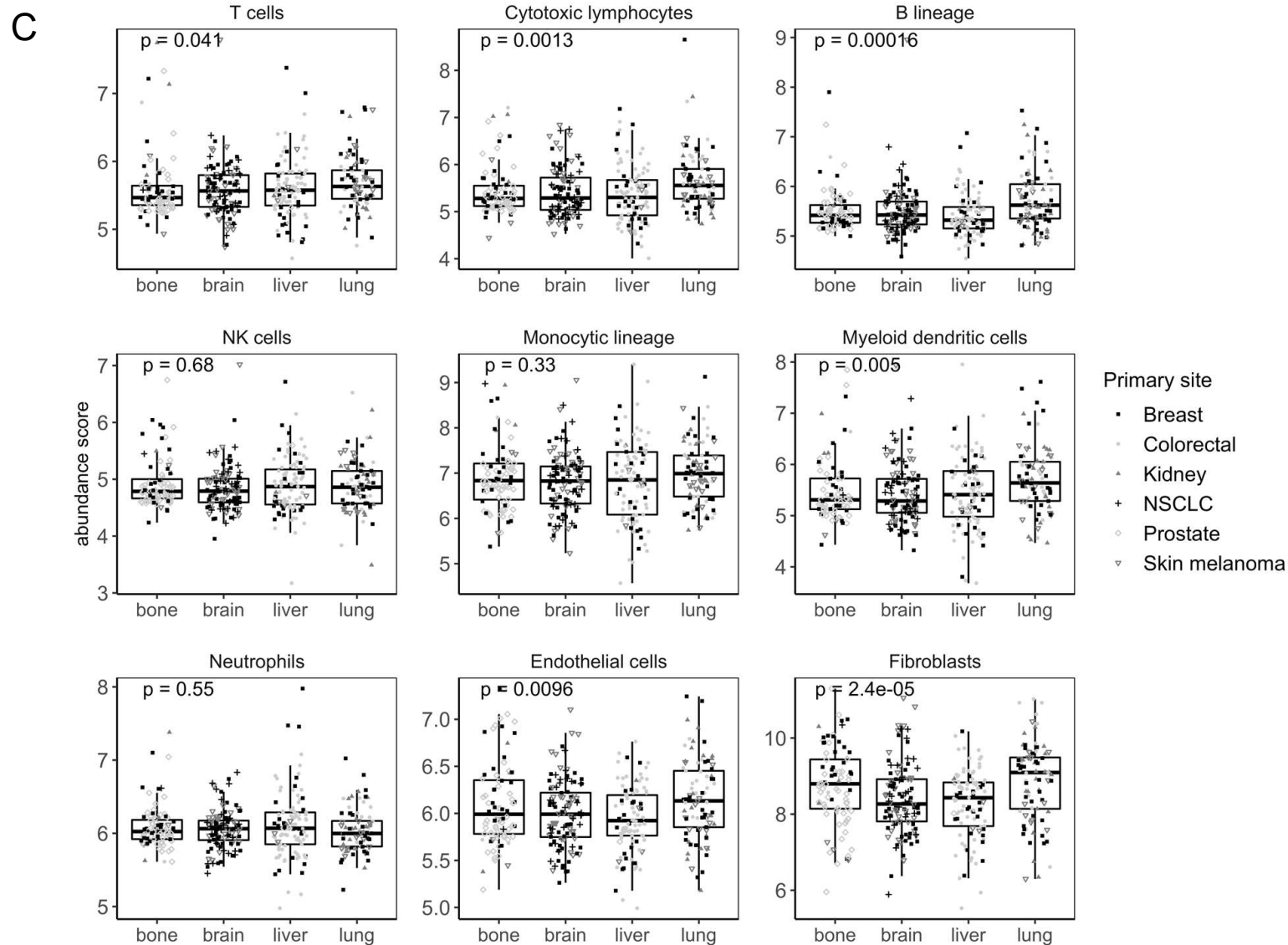
## Supplementary Figure 4

A



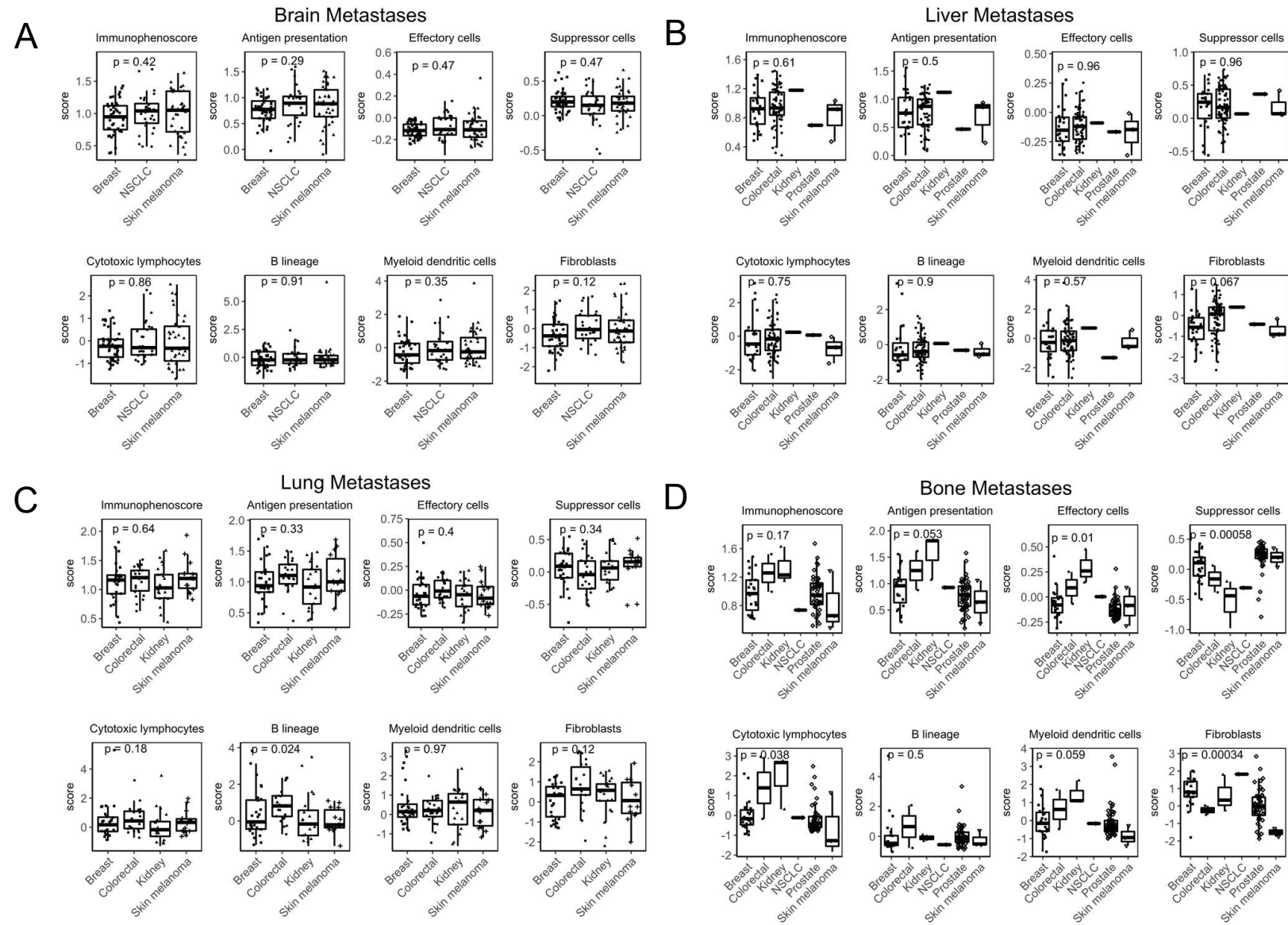
B





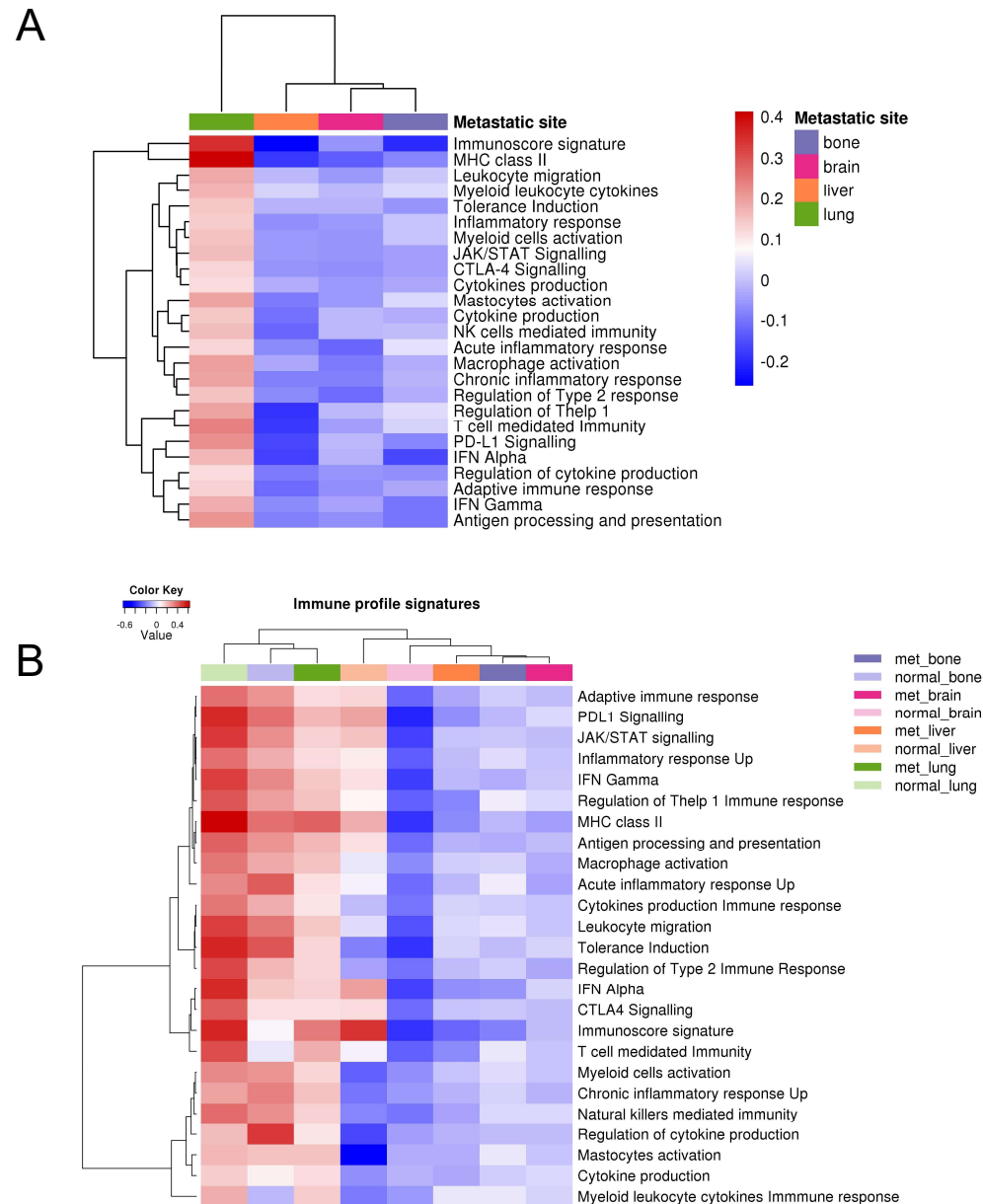
**Supplementary Figure 4. A.** Correlation between immunophenoscore (IPS) aggregated z-score and lung healthy tissue signature in lung metastatic samples. Results on IPS (B) and immune infiltration (C) on metastatic samples adjusting by healthy tissue signature score. NSCLC: non-small cell lung cancer.

## Supplementary Figure 5



**Supplementary Figure 5.** Immune scores in stratifying by primary site of origin in brain (A), liver (B), lung (C) and bone (D) metastasis. NSCLC: non-small cell lung cancer.

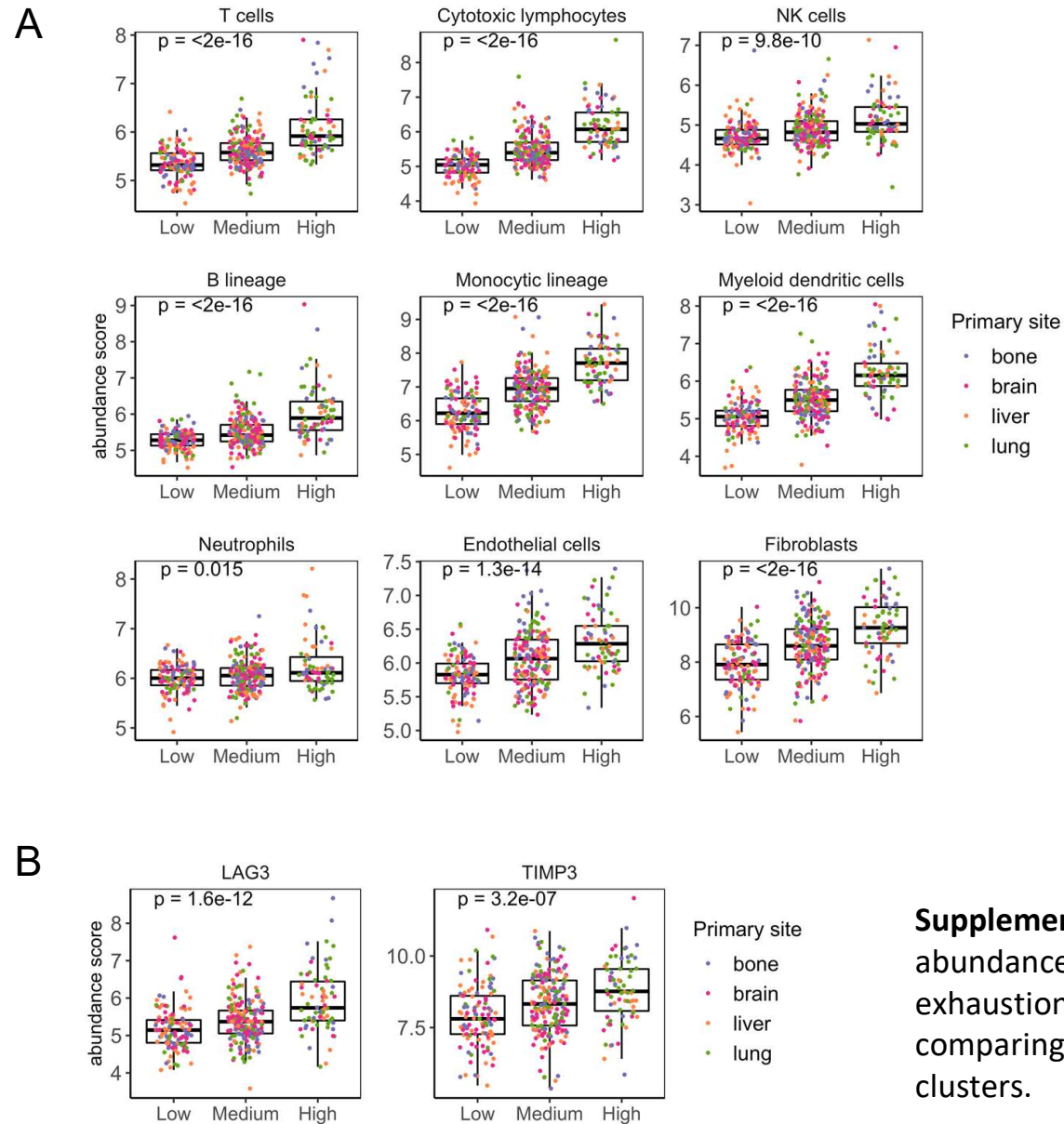
## Supplementary Figure 6



**Supplementary Figure 6. A.** Median values of GSVAs scores for each metastatic location. **B.** Heatmap plot of median GSVAs enrichment scores for the four metastatic sites of study and paired healthy tissues.



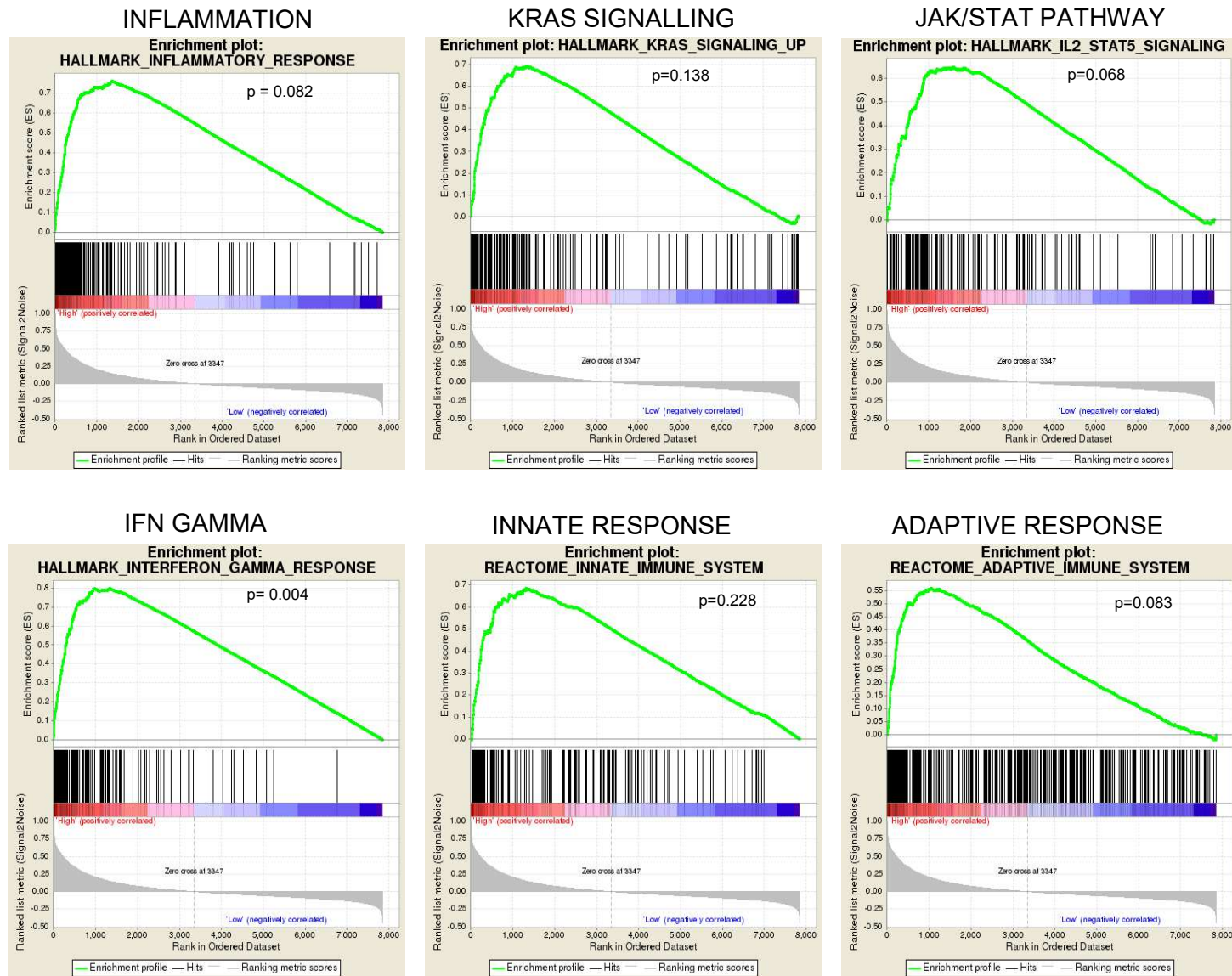
## Supplementary Figure 7



**Supplementary Figure 7.** Cell lineage abundance (A) and gene expression levels of exhaustion marker LAG3 and TIMP3 (B) comparing samples within the three immune clusters.

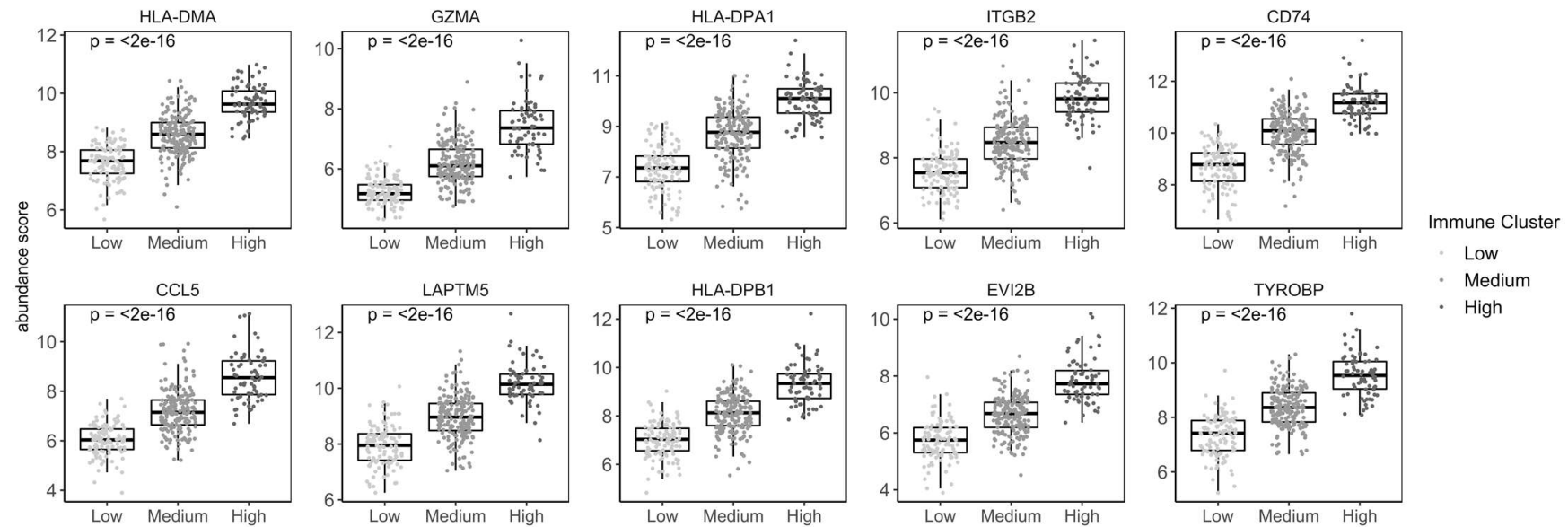


## Supplementary Figure 8



**Supplementary Figure 8.** Enrichment plots from Gene Set Enrichment Analysis (GSEA) performed between the HIC and the LIC metastatic samples. P-values indicate the FWER p-values for the enrichment scores.

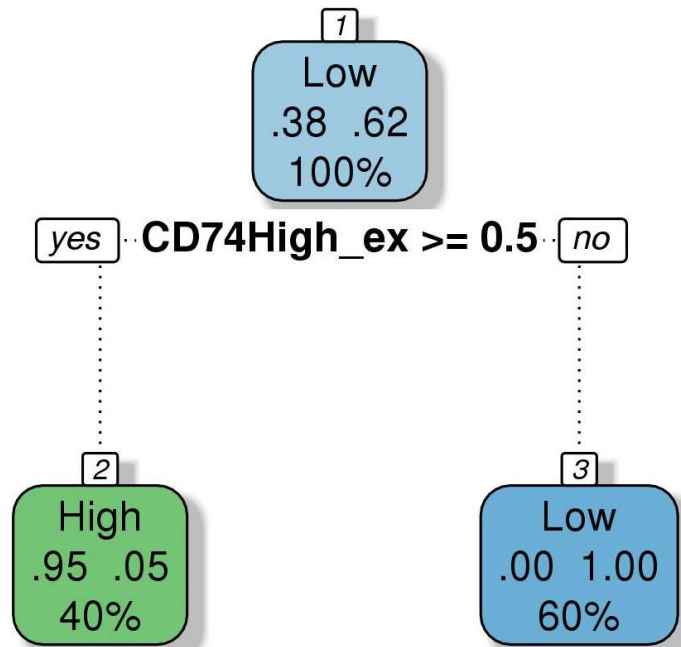
## Supplementary Figure 9



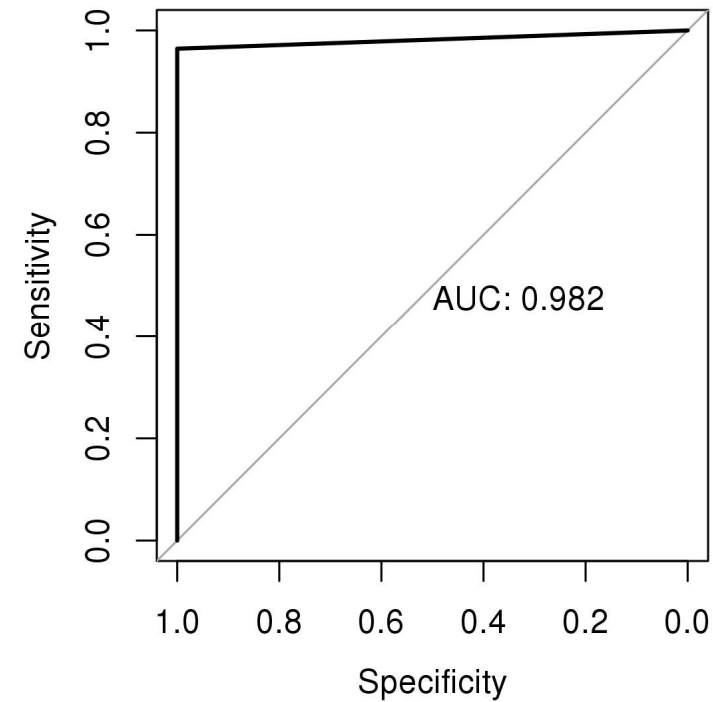
**Supplementary Figure 9.** Boxplots of the 10 most differentially expressed genes (DEGs) between samples belonging to HIC (n=70) and LIC (n=131) ImmunoClusters. Samples are compared by ImmuneCluster.

## Supplementary Figure 10

A



B



**Supplementary Figure 10. A.** Decision tree algorithm with 5-fold cross validation to discriminate between HIC and LIC samples in the training set.  $CD74$  gene was selected to be the best model, classifying correctly 100% of LIC samples and 95% of HIC samples. **B.** ROC curve of the predictive model based on  $CD74$  gene expression on testing data.