Figure S1. One-way logistic regression analysis of significantly differentially expressed m5C genes. Control vs. IPF. *P<0.05, **P<0.01, ***P<0.001.

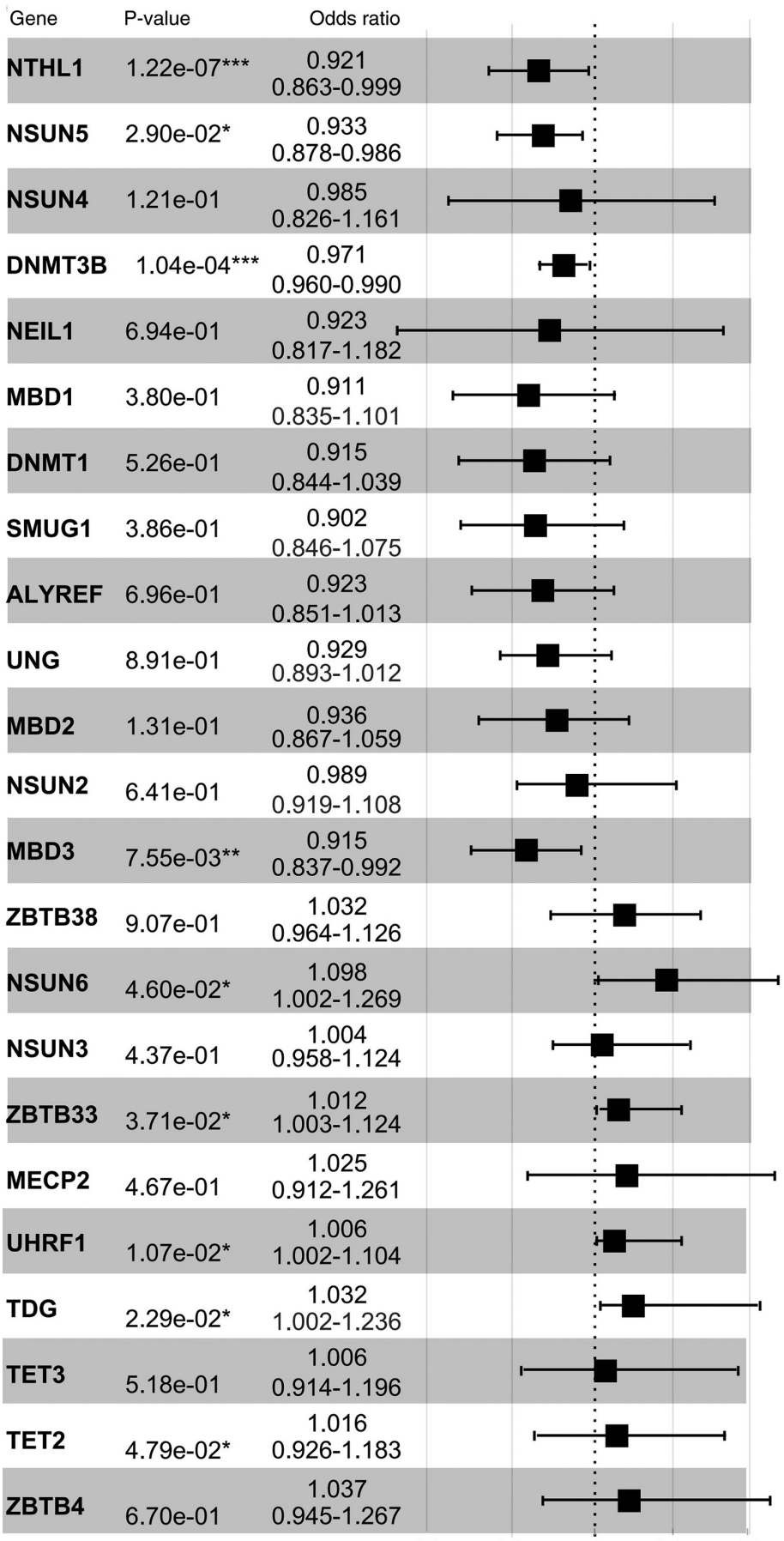

Figure S2. Plots of LASSO screening parameters. (A) Mean squared error. (B) Gene coefficient.

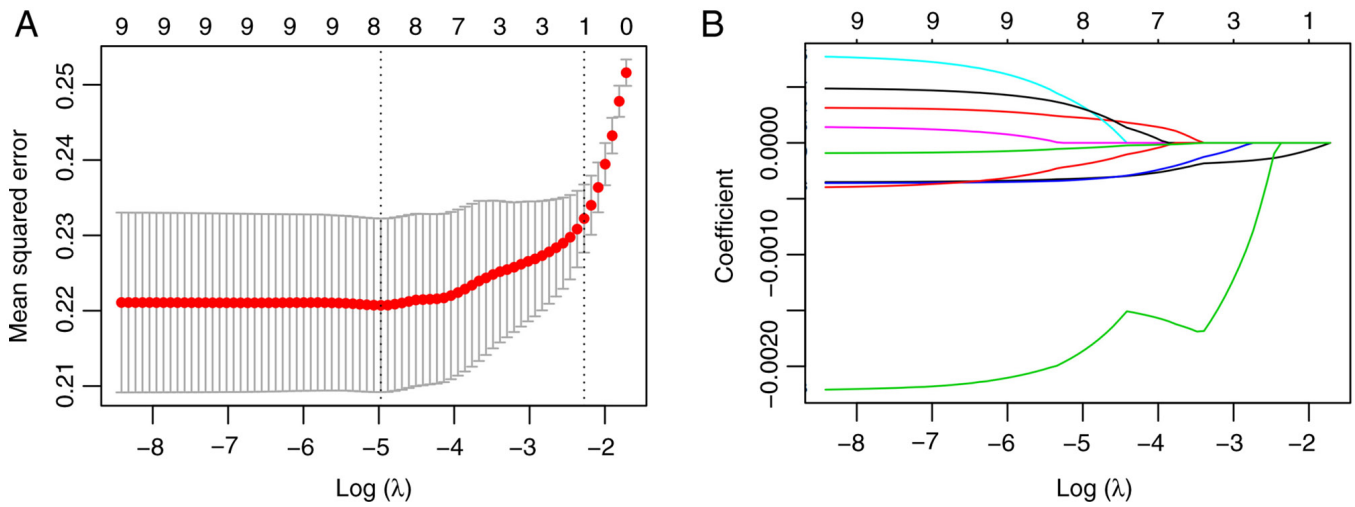

Figure S3. Receiver operating characteristic curves based on eight m5C genes in conjunction with the specificity and sensitivity parameters. (A) ROC curve of a diagnostic model based on 8 m5C genes in the training dataset. (B) Heat map of expression levels of 5-methylcytosine genes in the training set. (C) ROC curve of a diagnostic model based on 8 m5C genes in the independently validated dataset GSE124685. (D) Heat map of expression levels of 5-methylcytosine genes in the independently validated dataset GSE124685. IPF, idiopathic pulmonary fibrosis; CTRL, control; m5C, 5-methylcytosine; AUC, area under the curve.

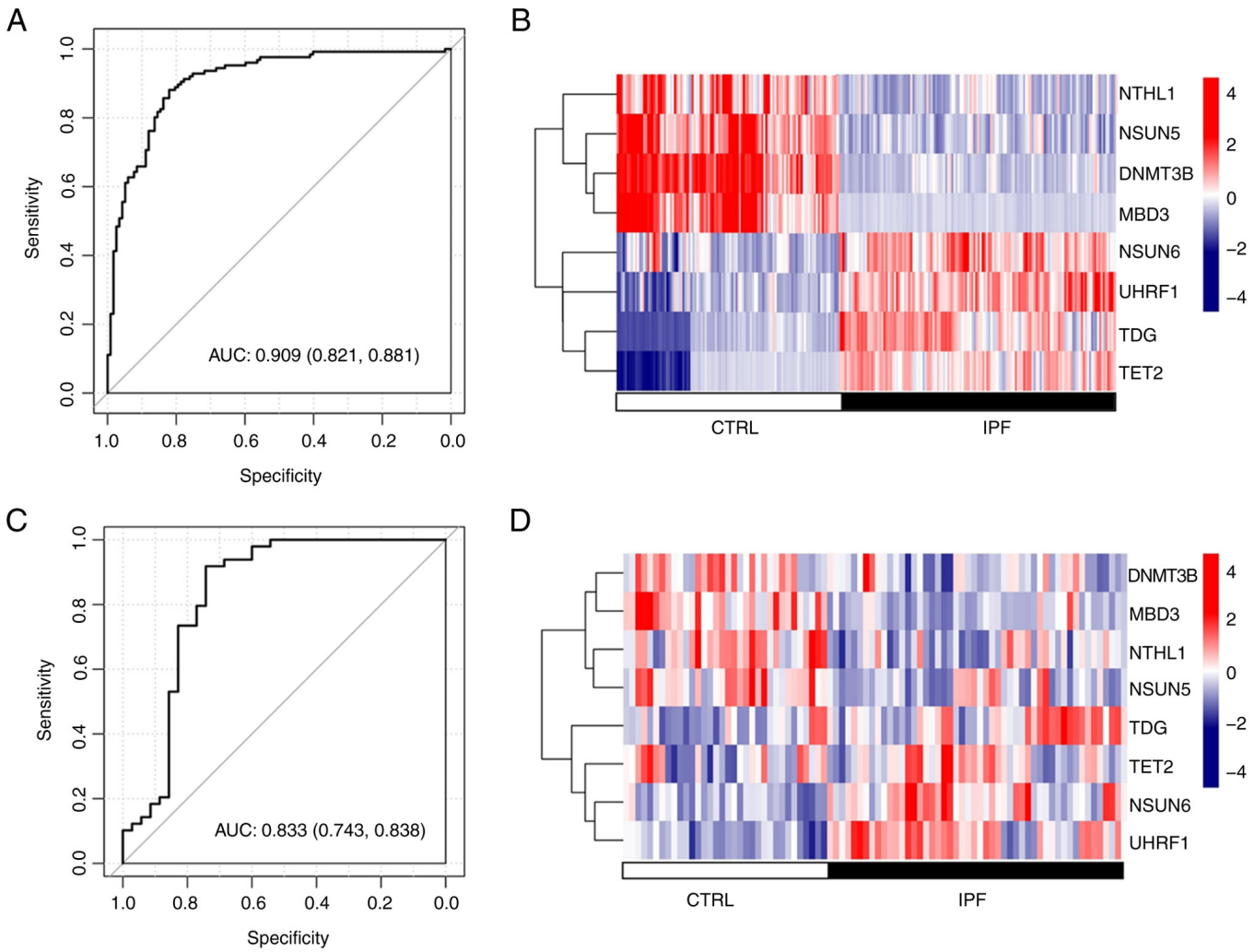

Figure S4. Differential analysis of subtype. Distribution of (A) immune cell type and (B) ESTIMATE score. *P<0.05, ***P<0.001.

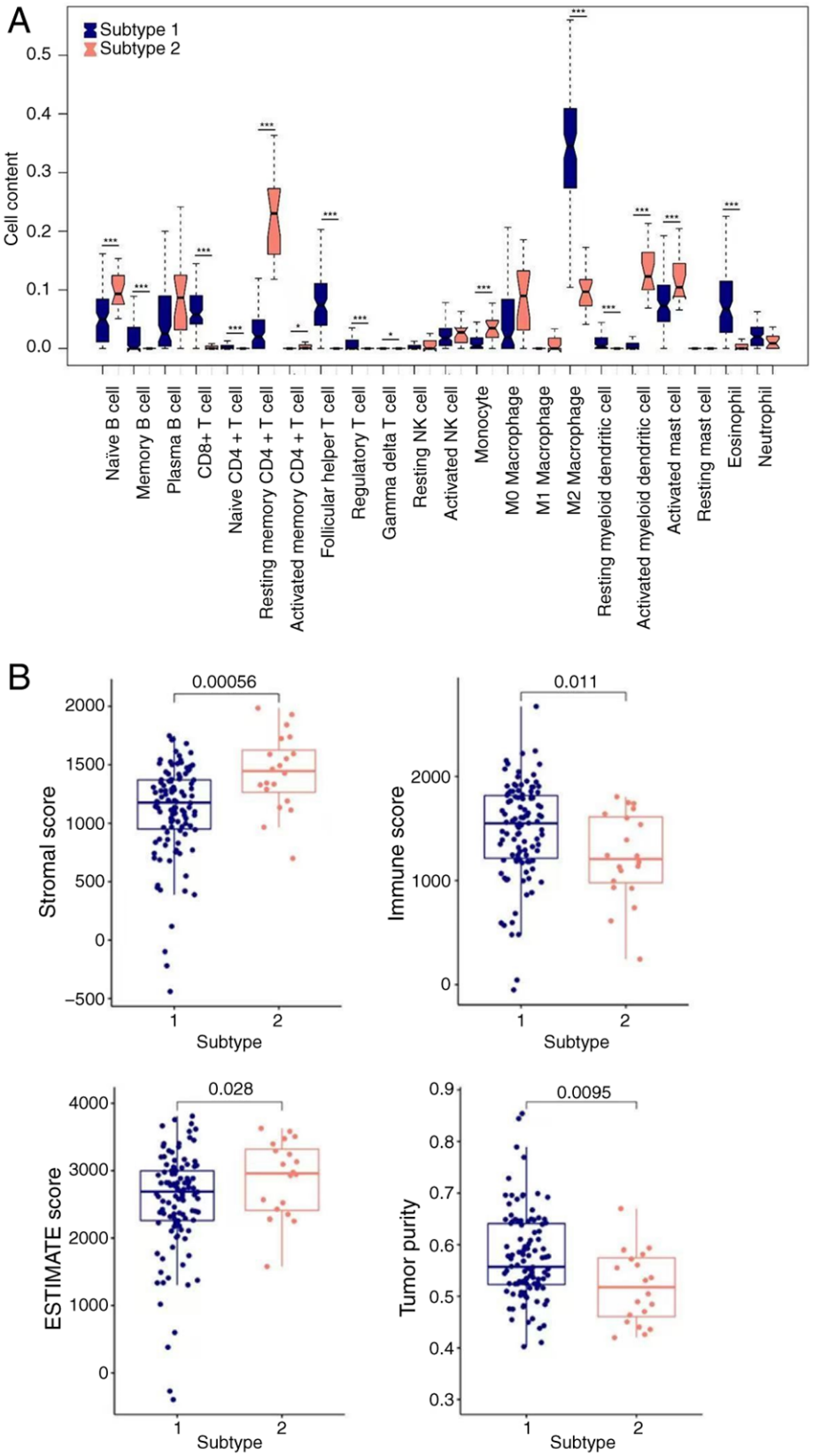

Figure S5. HLA family and immune checkpoint gene expression analysis of subtypes. Distribution of (A) expression levels of HLA family and (B) immune checkpoint genes. *P<0.05, ***P<0.001. HLA, human leukocyte antigen.

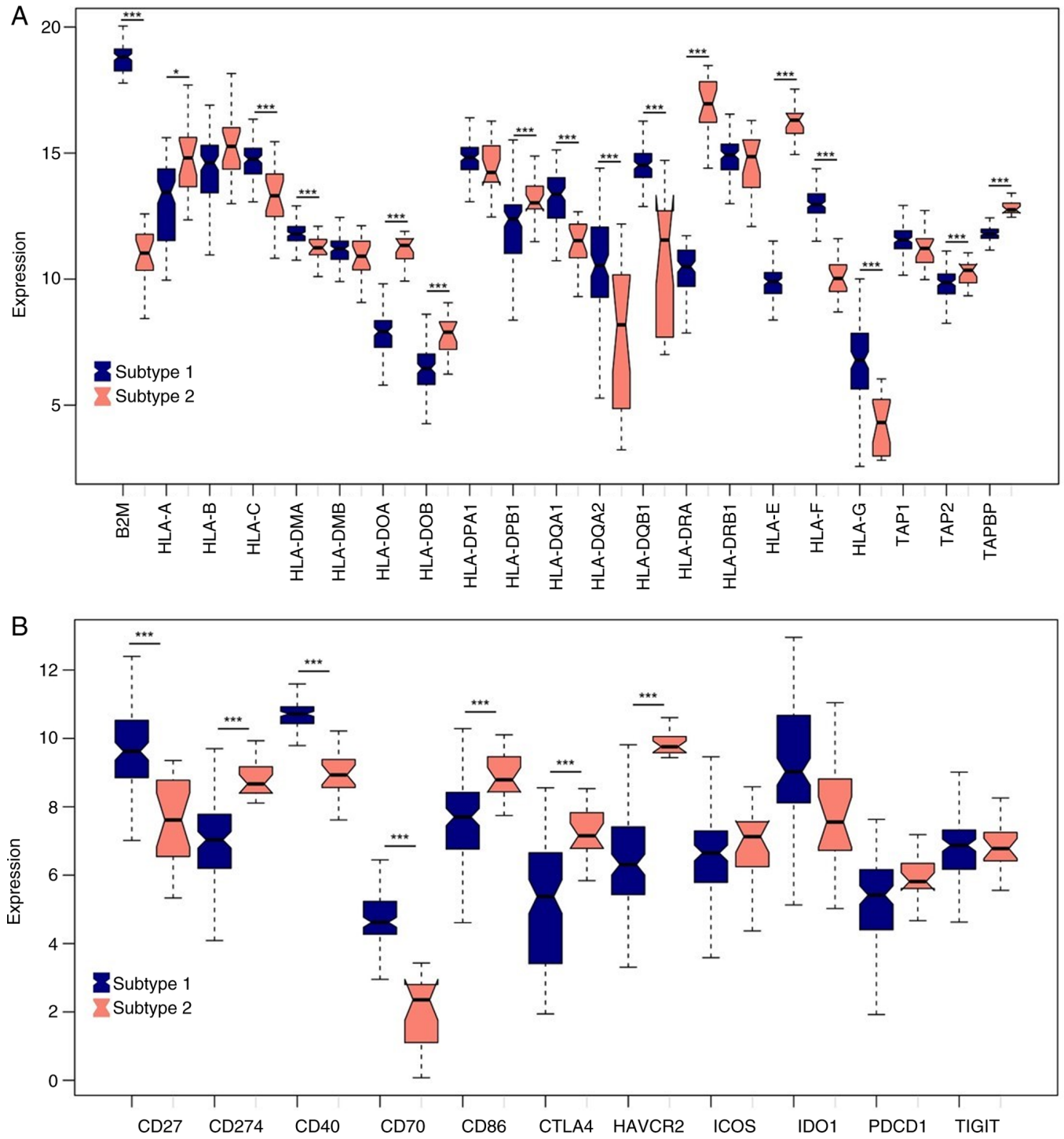

Figure S6. Subtype grouping of KEGG signaling pathways associated with subtype 1 (high m5C scores. KEGG, Kyoto Encyclopedia of Genes and Genomes.