AN IMMUNE-RELATED GENE PROGNOSTIC INDEX FOR HEAD AND NECK SQUAMOUS CELL CARCINOMA

SUPPLEMENTARY FIGURES

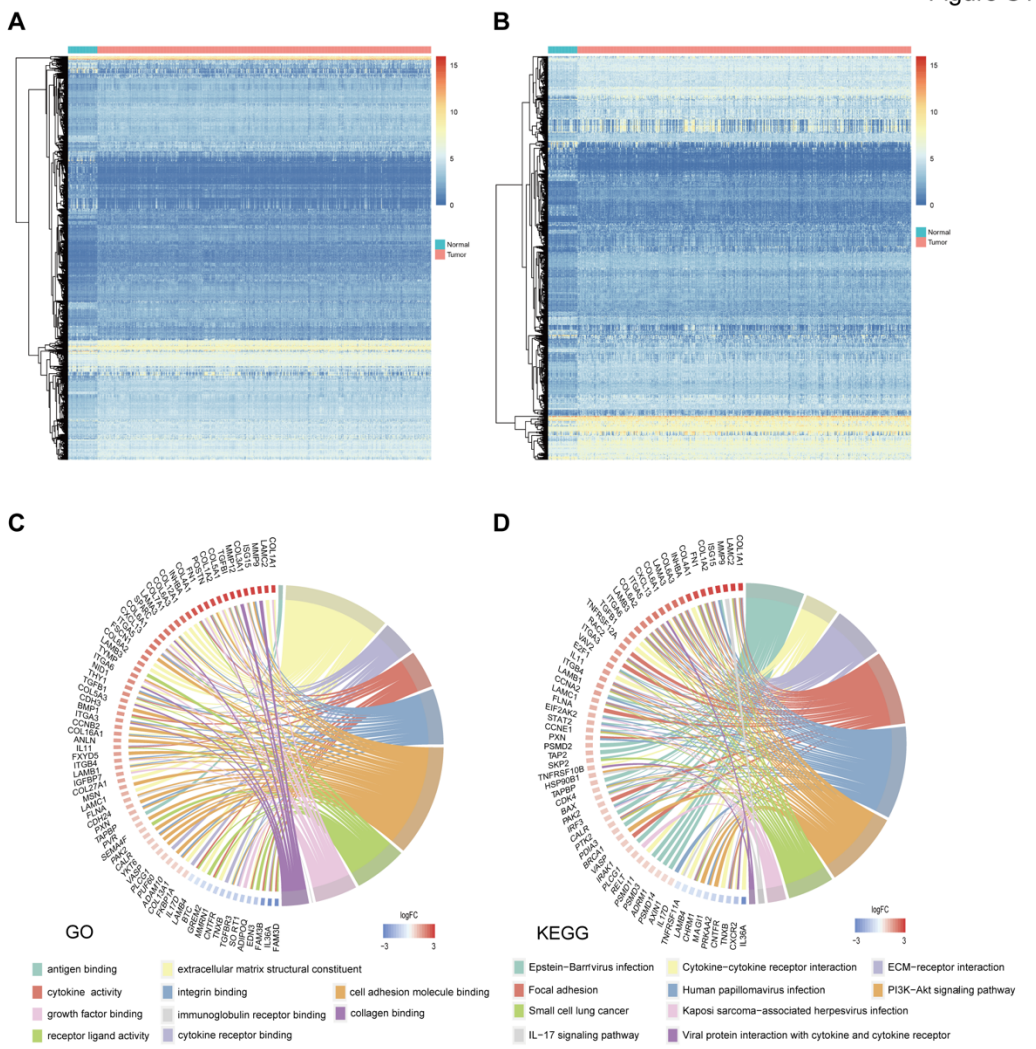

Figure S1. Differentially expressed immune-related genes in HNSCC.

(A) Heatmap displaying all differentially expressed genes (DEGs) between 502 HNSCC samples (red) and 44 para-cancer samples (blue) ($p < 0.05$, $|\log2FC| > 0.585$). **(B)** Heatmap displaying immune-related DEGs between 502 HNSCC samples (red) and 44 para-cancer samples (blue). **(C)** Gene Ontology (GO) enrichment analysis of the immune-related DEGs ($p < 0.05$). **(D)** Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis of the immune-related DEGs ($p < 0.05$).

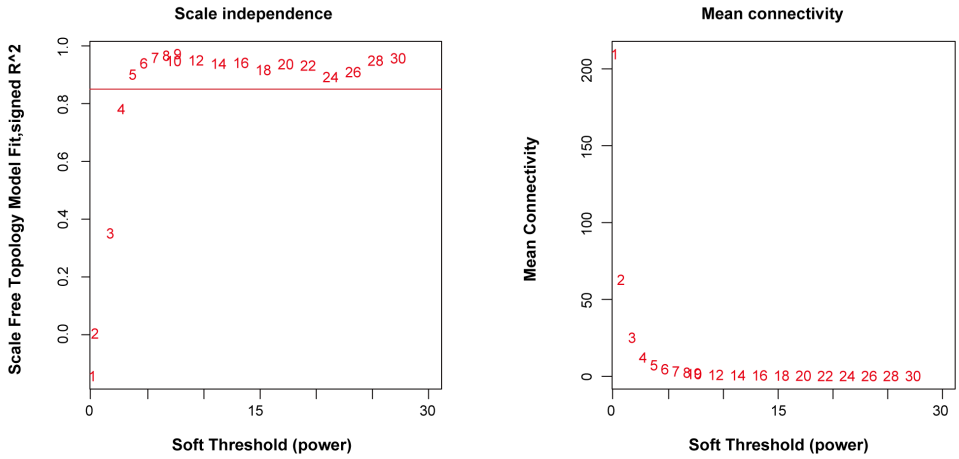

Figure S2. Determination of the soft-thresholding power in the WGCNA analysis.

In the left graph, the horizontal line indicates that the threshold value is 0.85. As seen from the graph, the optimal soft threshold for WGCNA was 5.

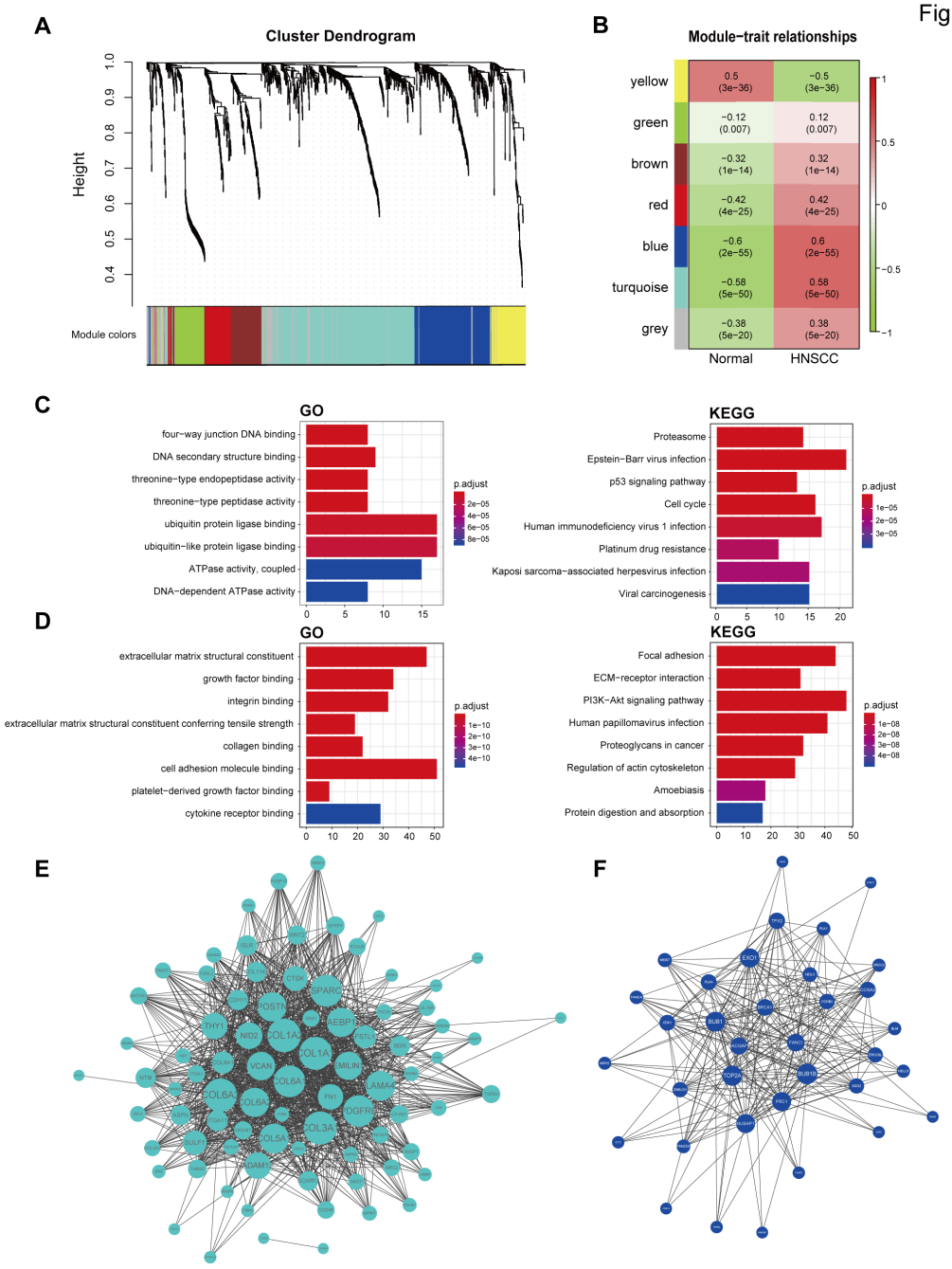

Figure S3. Identification of immune-related hub genes.

(A) Weighted gene coexpression network analysis (WGCNA) of immune-related differentially expressed genes with a soft threshold β = 5. **(B)** Gene modules related to HNSCC obtained by WGCNA. **(C)** Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways enriched in the genes of the blue module ($p < 0.05$). **(D)** GO and KEGG pathways enriched in the genes of the turquoise module ($p < 0.05$). **(E)** The network of the genes in the blue

module (weight of edge > 0.2). **(F)** The network of the genes in the turquoise module (weight of edge > 0.2).

Figure S4

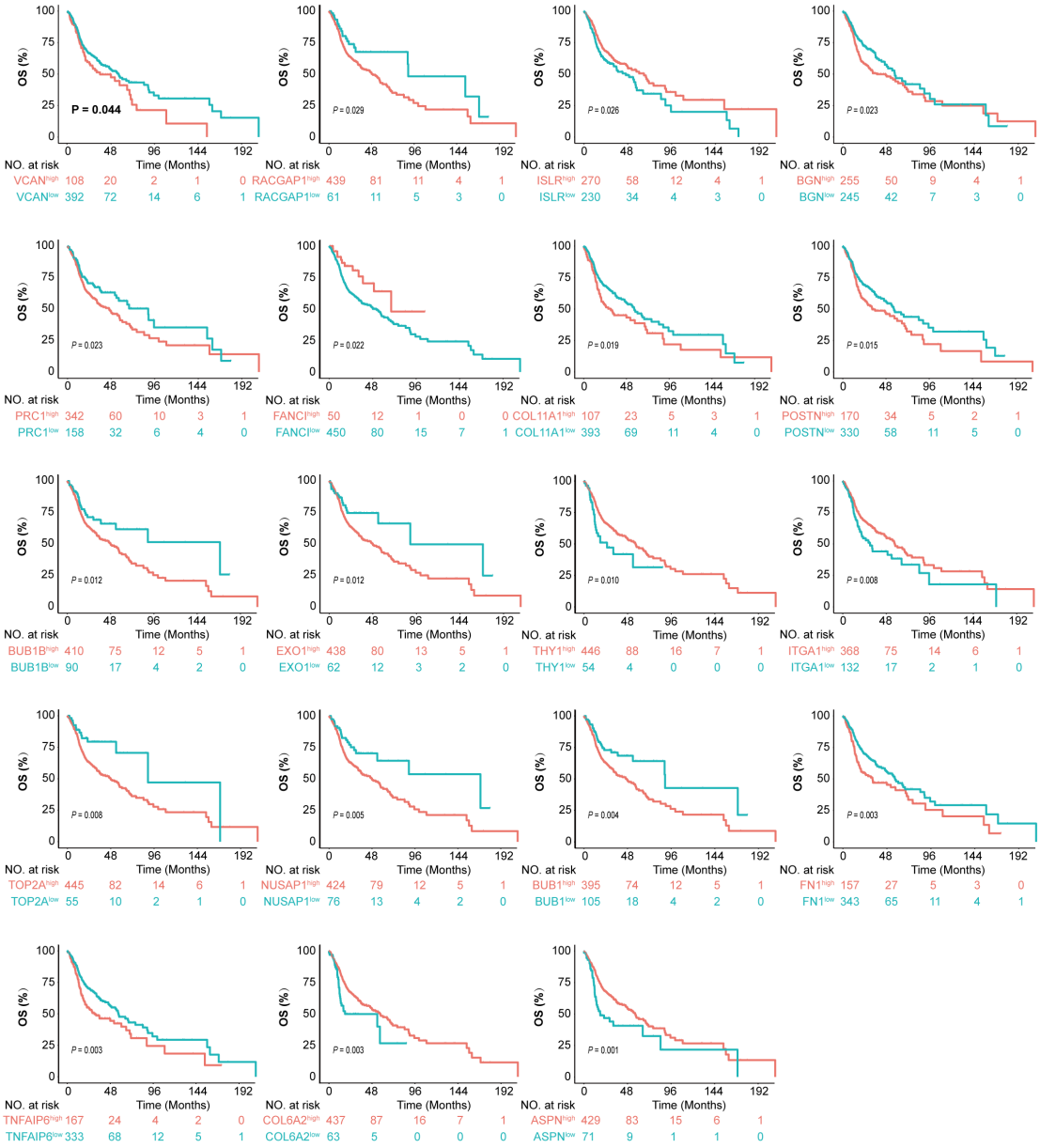

Figure S4. Kaplan-Meier curves of 22 immune-related hub genes.

Kaplan-Meier survival analysis of 22 immune-related genes in TCGA cohort.

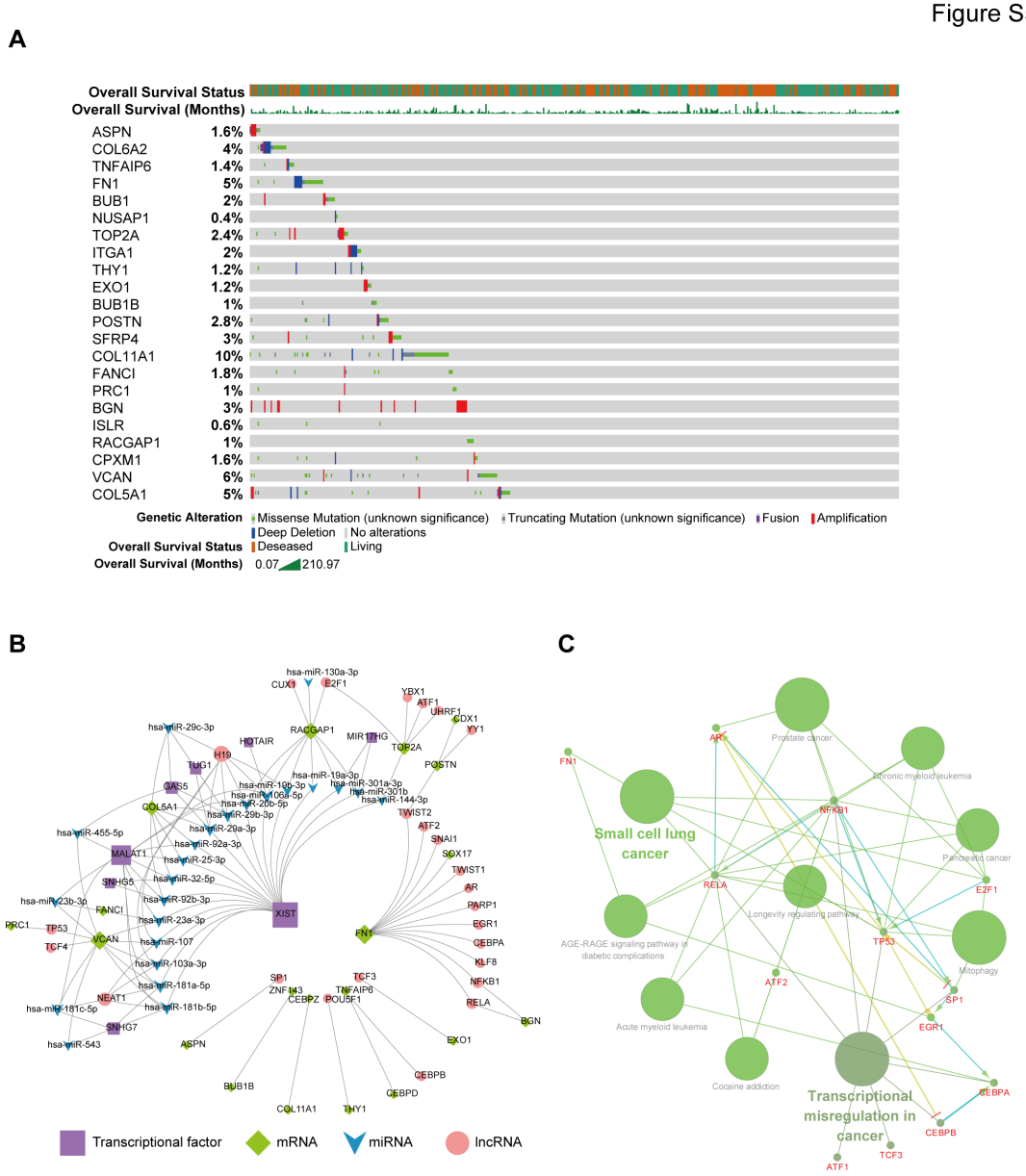

Figure S5. Molecular characteristics of immune-related hub genes.

(A) Alteration events for 22 immune-related hub genes are displayed. The upper bar plot indicates overall survival status in terms of overall survival time (months) per patient, whereas the left bar plot shows the mutation frequency of each gene. **(B)** The network summarizes complex connections between transcription factors (purple squares), mRNAs (green diamonds), miRNAs (blue triangles), and lncRNAs (red

circles). The size of the node is positively correlated with the degree of the node. **(C)** KEGG pathways (circles) enriched in the genes (red) of the transcription factor and ncRNA network by the ClueGO plugin of Cytoscape ($p < 0.05$). The size of the circle indicates the number of genes in the enrichment pathway, the color of the circle indicates the approximation among different pathways, and links indicate genes in the enrichment pathway.

Figure S6

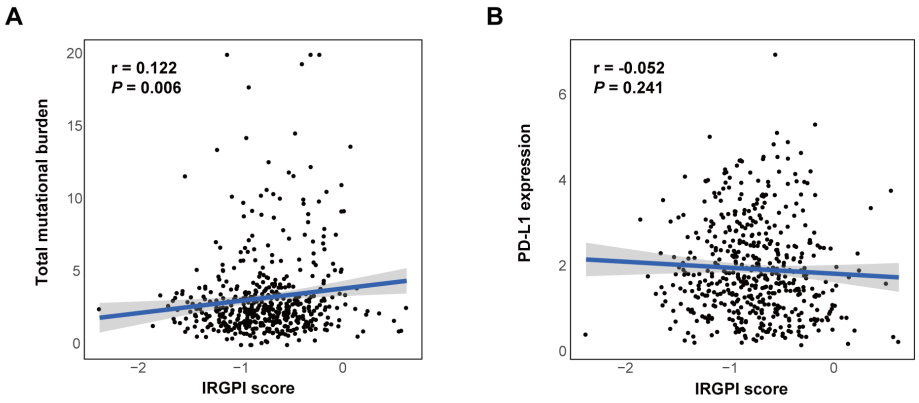

Figure S6. The relationship between IRGPI and total mutational burden and PD-L1 expression.

(A) Correlation analysis between IRGPI and total mutational burden. **(B)** Correlation analysis between IRGPI and PD-L1 expression.

Figure S7

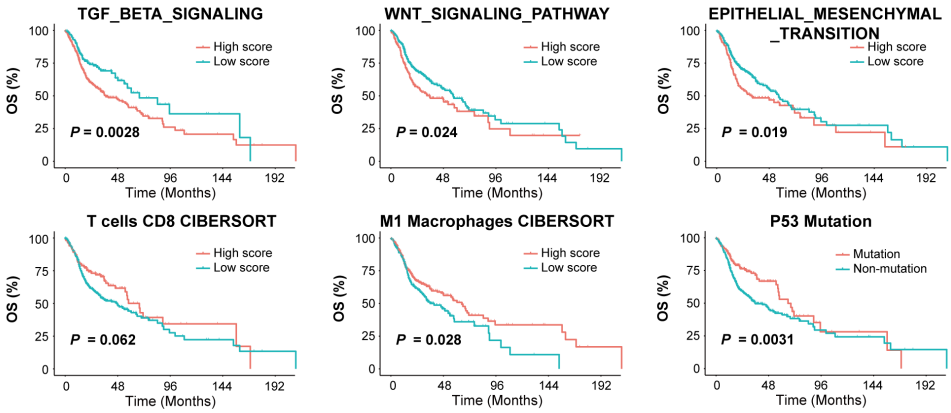

Figure S7. Kaplan-Meier curves of certain molecular and immune-related gene sets.

Kaplan-Meier survival analysis of certain signal pathways, the proportion of different immune cells and TP53 mutation in TCGA cohort.

Figure S8

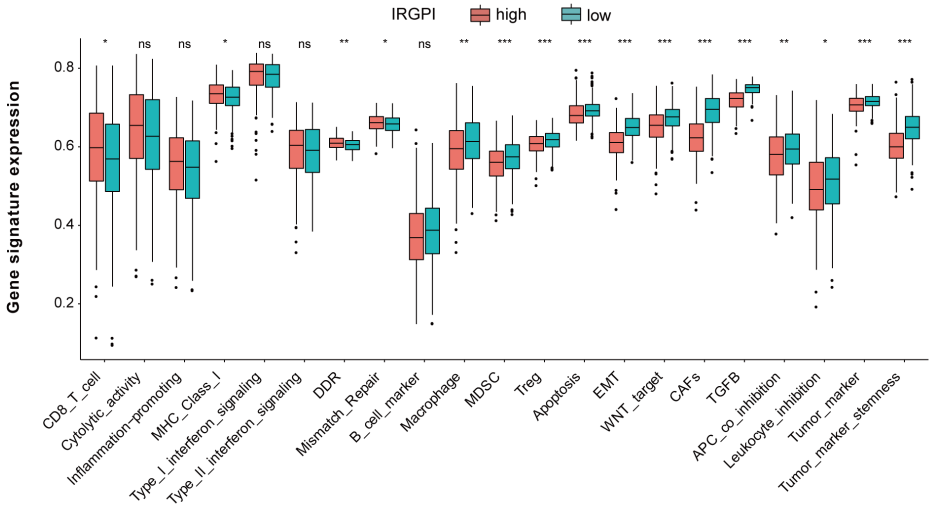

Figure S8. The molecular and immune-related function of different IRGPI subgroup.

The gene sets of molecular and immune-related function were analyzed by the single simple gene set enrichment analysis (ssGSEA) and then compared between different IRGPI subgroups. The scattered dots represent the ssGSEA scores of the two subgroups. The thick lines represent the median value. The bottom and top of the boxes are the 25th and 75th percentiles (interquartile range), respectively. Significant statistical differences between the two subgroups were assessed using the Wilcoxon test (ns: not significant, *$p < 0.05$, ** $p < 0.01$, *** $p < 0.001$).