

Identification of a ubiquitin related gene signature for predicting prognosis and immune infiltration in breast cancer based on single cell and bulk RNA sequencing analysis

Zhengbo Fang

Second Affiliated Hospital of Harbin Medical University

Jiaqi Li

Second Affiliated Hospital of Harbin Medical University

Man Li

Harbin Medical University Cancer Hospital

Fei Ma

Second Affiliated Hospital of Harbin Medical University

Yansong Liu

Second Affiliated Hospital of Harbin Medical University

Weilun Cheng

Second Affiliated Hospital of Harbin Medical University

Yunqiang Duan

Second Affiliated Hospital of Harbin Medical University

Anbang Hu

Second Affiliated Hospital of Harbin Medical University

Jiarui Zhang

Second Affiliated Hospital of Harbin Medical University

Mingcui Li

Second Affiliated Hospital of Harbin Medical University

Yanling Li

Second Affiliated Hospital of Harbin Medical University

Yuhang Shang

Second Affiliated Hospital of Harbin Medical University

Jiangwei Liu

Second Affiliated Hospital of Harbin Medical University

Yuanhao Ji

Second Affiliated Hospital of Harbin Medical University

Shan Gao

Second Affiliated Hospital of Harbin Medical University

Yiyun Zhang

3253@hrbmu.edu.cn

Harbin Medical University Cancer Hospital

Research Article

Keywords: breast cancer, ubiquitination, prognostic signature, immune microenvironment, biomarker, SORBS1

Posted Date: March 27th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-8886340/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Additional Declarations: No competing interests reported.

Title: Identification of a ubiquitin related gene signature for predicting prognosis and immune infiltration in breast cancer based on single cell and bulk RNA sequencing analysis

Running Title: Ubiquitin Signature in Breast Cancer

Zhengbo Fang^{1†}, Jiaqi Li^{1†}, Man Li^{2†}, Fei Ma¹, Yansong Liu¹, Weilun Cheng¹, Yunqiang Duan¹, Anbang Hu¹, Jiarui Zhang¹, Mingcui Li¹, Yanling Li¹, Yuhang Shang¹, Jiangwei Liu¹, Yuanhao Ji¹, Shan Gao^{3*}, Yiyun Zhang^{2*}

† These authors contributed equally to this work.

* Corresponding authors.

1 Department of General Surgery, The Second Affiliated Hospital of Harbin Medical University, Harbin, China.

2 Department of Endoscope, Harbin Medical University Cancer Hospital, Harbin, China.

3 Department of Pathology, The Second Affiliated Hospital of Harbin Medical University, Harbin, China.

* Correspondence

Yiyun Zhang

Department of Endoscope, Harbin Medical University Cancer Hospital, Harbin, China.

Email: 3253@hrbmu.edu.cn

Shan Gao

Department of Pathology, The Second Affiliated Hospital of Harbin Medical University, Harbin, China.

Email: shangao@hrbmu.edu.cn

Abstract

Background

Ubiquitination drives breast cancer (BC) progression, yet the clinical value of ubiquitin-related genes (UbRGs) remains underexplored. This study aimed to develop a UbRG-based prognostic model and investigate its associations with the tumor immune microenvironment (TIME) and therapeutic responses.

Methods

Transcriptomic and clinical data from the TCGA-BRCA cohort were analyzed to identify differentially expressed UbRGs. A ubiquitin-related risk score (URS) was constructed using LASSO and Cox regression. Single-cell RNA sequencing (scRNA-seq) data (GSE161529) were integrated via the Scissor algorithm to identify risk-associated cell populations. Furthermore, weighted gene co-expression network analysis (WGCNA) combined with machine learning approaches was applied to identify a core hub gene, which was subsequently validated using immunohistochemistry (IHC) in a clinical cohort.

Results

A six-gene URS independently stratified patients into high- and low-risk groups with significantly different overall survival. High-risk tumors exhibited an immunosuppressive TIME, featuring decreased lymphocyte infiltration and M2 macrophage enrichment, which was corroborated at the single-cell level. Conversely, low-risk patients demonstrated predicted sensitivity to targeted therapies, including CDK4/6 and PARP inhibitors. Through multi-omics and machine learning integration, SORBS1 was identified as the critical hub gene. IHC validation confirmed that SORBS1 protein expression was significantly downregulated in BC tissues and correlated with improved survival.

Conclusions

The scRNA-seq and machine learning-integrated URS provides robust prognostic stratification and therapeutic guidance in BC. Our findings link ubiquitination to immune exclusion and establish SORBS1 as a clinically validated, protective biomarker.

Keywords

breast cancer; ubiquitination; prognostic signature; immune microenvironment; biomarker; SORBS1

1. Introduction

Breast cancer remains a major cause of cancer-related death among women worldwide and continues to pose a significant clinical challenge, even with advances in treatment (Kim et al. 2025). Targeted therapies, including CDK4/6 inhibitors, PARP inhibitors, and antibody-drug conjugates, have clearly improved survival for many patients (Bai et al. 2025; Sabit et al. 2025). However, clinical outcomes vary widely. A considerable proportion of patients eventually develops resistance to these treatments, resulting in disease progression or relapse (Gao et al. 2025). Such differences in treatment response are driven by complex molecular changes within tumor cells that are not fully captured by routine clinicopathological indicators, such as TNM stage or receptor status (Saha et al. 2025). This gap underscores the need for biomarkers that more accurately reflect the molecular processes underlying tumor aggressiveness and treatment resistance.

Ubiquitination is a common post-translational modification and plays an essential role in regulating these intracellular processes (Zhu et al. 2025). In addition to its well-known function in protein degradation, the ubiquitin system, coordinated by E1, E2, and E3 enzymes, controls the stability and activity of many proteins involved in cancer-related pathways, including the PI3K/AKT/mTOR signaling pathway and DNA damage repair (Koo et al. 2023). Growing evidence also indicates that

abnormal ubiquitination influences the tumor immune microenvironment (TIME) (Huang et al. 2025).

Certain E3 ligases have been reported to regulate the turnover of immune checkpoint molecules, such as PD-L1 (Lemma et al. 2023), as well as antigen-presenting molecules, such as MHC class I (Dhatchinamoorthy et al. 2021), thereby shaping how tumors interact with the immune system. Despite these insights, the clinical application of ubiquitin-related mechanisms remains limited (Spano and Catara 2023a). Most studies focus on individual genes, which do not adequately reflect the broader relationship between the ubiquitin network and immune infiltration, as well as genomic instability, in patient tumors (Ma et al. 2025).

Another limitation in current prognostic research is the heavy reliance on bulk transcriptomic data, which can mask signals from specific cell types (Li et al. 2025). Ubiquitination activity differs between tumor cells and cells in surrounding immune or stromal compartments, underscoring the importance of cell-type resolution for accurate risk assessment (Huang et al. 2025). At the same time, although large-scale analyses can identify candidate genes, validation at the protein level in human tissues is often lacking (Diamandis 2012). SORBS1, a scaffold protein involved in cytoskeletal organization and insulin-related signaling, has been suggested to act as a tumor suppressor (Jin et al. 2022). However, its prognostic value and its potential link to ubiquitin-related immune regulation in breast cancer have not been clearly defined (Feng et al. 2024).

To address these issues, we developed a multi-omics framework to assess the prognostic and predictive relevance of ubiquitin-related genes (UbRGs) in breast cancer. Using the TCGA-BRCA dataset, we constructed a six-gene risk signature and examined its association with immune evasion, tumor mutational burden, and drug sensitivity. To better understand the cellular sources of risk-related signals, we incorporated single-cell RNA sequencing data and applied the Scissor algorithm. In addition, the

key gene SORBS1 was validated at the protein level using immunohistochemistry in an independent cohort of clinical samples. Together, this study aims to provide a practical tool for risk stratification and to elucidate the immunogenomic features that link ubiquitination to breast cancer progression.

2. Methods

2.1. Data Acquisition and Preprocessing

We retrieved mRNA expression data and clinical information from The Cancer Genome Atlas (TCGA-BRCA). The study population was strictly filtered to exclude patients with missing survival data or an overall survival (OS) of less than 30 days, thereby minimizing bias. The final cohort comprised 847 tumor samples and 99 normal tissue controls. For model construction and internal validation, tumor samples were randomly split into a training set (70%, $n = 593$) and a testing set (30%, $n = 254$). A summarized comparison of clinicopathological features is provided in Table 1. Simultaneously, a total of 1,131 ubiquitin-related genes (UbRGs) were retrieved from the iUUCD 2.0 database (Zhou et al. 2018). To facilitate single-cell-resolution analysis, the GSE161529 dataset, containing single-cell transcriptomes from three primary breast tumors (representing ER+, HER2+, and Triple-Negative subtypes), was obtained from the Gene Expression Omnibus (GEO) (Wu et al. 2021).

2.2. Construction of the Ubiquitin-Related Prognostic Signature

Differential expression analysis between tumor and normal tissues was conducted using the edgeR package (Robinson et al. 2010). Ubiquitin-related differentially expressed genes (Ub-DEGs) were identified based on a threshold of $|\log_2 \text{fold change}| > 1$ and $p < 0.05$. To establish a robust prognostic model, we first screened Ub-DEGs using univariate Cox regression in the training cohort. Genes significantly associated with OS ($p < 0.05$) were subsequently entered into a Least Absolute Shrinkage

and Selection Operator (LASSO) Cox regression model using the glmnet package (Simon et al. 2011).

Optimal parameter selection was determined via 10-fold cross-validation. The Ubiquitin-Related Risk Score (URS) for each patient was calculated using the following formula:

$$URS = \sum_{i=1}^n Coef_i * Exp_i$$

Patients were stratified into high- and low-risk groups based on the median URS. Survival differences were evaluated via Kaplan-Meier curves and log-rank tests. The model's predictive accuracy was quantified using time-dependent Receiver Operating Characteristic (ROC) curves in both the training and testing cohorts.

2.3. Clinical Validation and Nomogram Development

To evaluate whether the URS is independent of clinical variables (such as age, clinical stage, and TNM classification), we performed both univariate and multivariate Cox regression analyses. The URS, age, clinical stage, T stage, and N stage were included in the predictive nomogram, which was constructed using the rms package. The performance of the nomogram was evaluated using calibration plots, which compared the predicted survival probabilities at 1, 7, and 10 years with the actual observed survival rates.

2.4. Functional Enrichment Analysis

To elucidate the biological mechanisms underlying the risk stratification, we identified DEGs between the high- and low-risk groups ($|\log_2 \text{ fold change}| > 1$, $p < 0.05$). Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analyses were performed using the clusterProfiler package to identify significantly enriched biological processes and signaling pathways (Yu et al. 2012).

2.5. Immunogenomic Profiling and Therapeutic Prediction

We employed a multi-faceted approach to characterize the tumor microenvironment (TME):

Immune Infiltration: The abundance of 22 immune cell types was estimated using the CIBERSORT algorithm (Newman et al. 2015). To further analyze immune cell signatures, Single-sample Gene Set Enrichment Analysis (ssGSEA) was applied to quantify 28 immune cell signatures (Wright et al. 2003). Immune score, ESTIMATE score, and tumor purity were calculated using the ESTIMATE package (Yoshihara et al. 2013). Differences between high- and low-risk groups were compared using the Wilcoxon rank-sum test.

Genomic Alterations: Somatic mutation data were processed using the maftools package to calculate Tumor Mutational Burden (TMB), defined as the total number of non-synonymous mutations per megabase (Mayakonda et al. 2018).

Drug Sensitivity: To evaluate therapeutic potential, we estimated the half-maximal inhibitory concentration (IC50) for chemotherapeutic and targeted agents using the oncoPredict package from the GDSC database (Maeser et al. 2021).

2.6. Single-Cell Resolution of Risk Phenotypes

Single-cell RNA sequencing data were processed using Seurat (v4.5.1), following a standard pipeline of quality control, normalization, dimensionality reduction, and clustering (Hao et al. 2021). Cell types were annotated referencing canonical markers and the CellMarker 2.0 database (Hu et al. 2023). To bridge bulk and single-cell data, we applied the Scissor (Single-Cell Identification of Subpopulations with Bulk Sample Phenotype Correlation) algorithm (Sun et al. 2022). This approach integrated the TCGA bulk expression profiles with the URS phenotype to identify specific single-cell subpopulations significantly associated with high- or low-risk status. Further sub-clustering of T cells and macrophages was performed to refine the characterization of the immune landscape.

2.7. Hub Gene Identification and Experimental Validation

A multi-step strategy was employed to identify a clinically relevant hub gene:

1. Screening: Weighted Gene Co-expression Network Analysis (WGCNA) was performed to identify gene modules highly correlated with the URS (Langfelder and Horvath 2008).
2. Ranking: The six genes from the prognostic signature were ranked using two machine learning algorithms: Random Forest and XGBoost. The gene consistently ranked highest across these methods, and WGCNA was designated as the hub gene.
3. Validation (IHC): Protein-level expression of the hub gene (SORBS1) was validated in a cohort of 136 breast tumor tissues and 46 normal tissues (Shanghai Outdo Biotech Co., Ltd.). Immunohistochemistry was performed using a rabbit polyclonal anti-SORBS1 antibody (Proteintech, 13854-1-AP; 1:200 dilution). Staining intensity was quantified via Average Optical Density (AOD) using Fiji ImageJ.

2.8. Statistical Analysis

Statistical analyses were performed using R version 4.4.2. Two-sided p-values < 0.05 were considered statistically significant unless otherwise specified.

3. Results

3.1. Identification of the Ubiquitin-Related Signature and Risk Stratification

Through differential expression analysis comparing breast cancer tissues with normal controls, we identified 5,007 differentially expressed genes (DEGs) (Figure 1A). By intersecting these with the ubiquitin-related gene set, we identified 148 ubiquitin-associated DEGs (Ub-DEGs) that potentially drive tumorigenesis (Figure 1B). Following univariate Cox screening and LASSO Cox regression

analysis (to minimize overfitting), a six-gene prognostic signature was established (Figure 1C-E). This signature integrates the expression levels of DCAF13, PPP2R2B, LONRF3, DTX1, SORBS1, and DCAF10. The Ubiquitin-Related Risk Score (URS) is calculated as follows: $URS = (0.375)*DCAF13 + (-1.204)*PPP2R2B + (0.330)*LONRF3 + (-0.150)*DTX1 + (-0.044)*SORBS1 + (-0.281)*DCAF10$.

Note: Coefficients are rounded to three decimal places for brevity.

All six genes showed significant expression differences between the high- and low-risk groups (Figure 1F), with expression trends consistent with those observed at the protein level in immunohistochemistry results from the Human Protein Atlas (HPA) (Figure 1G).

3.2. Prognostic Robustness and Clinical Independence

The URS proved effective at distinguishing patient outcomes (Figure 2A, B). In the training cohort, patients stratified into the high-risk group experienced significantly shorter overall survival compared to those in the low-risk group (Log-rank $p < 0.0001$) (Figure 2C). The model achieved time-dependent Area Under the Curve (AUC) values of 0.73, 0.70, and 0.72 for 1-, 7-, and 10-year survival, respectively (Figure 2D). We successfully validated these findings in both the testing cohort ($p = 0.027$) and the entire TCGA cohort ($p < 0.0001$), confirming the stability of the signature across datasets (Figure 2E-H).

To determine whether the URS provides prognostic value independently of standard clinical variables, we first conducted a univariate Cox regression analysis. This analysis identified age, clinical stage, T stage, N stage, M stage, and URS as significant predictors of survival (Figure 3A). In the subsequent multivariate analysis, the URS remained a strong independent prognostic factor (HR = 3.20; 95% CI: 2.14 – 4.79; $p < 0.0001$), along with age, clinical stage, and N stage (Figure 3B). Based on these results,

we constructed a nomogram integrating the URS, age, clinical stage, T stage, and N stage to improve individualized survival prediction (Figure 3C). Calibration plots for 1-, 7-, and 10-year survival showed excellent agreement between the predicted probabilities and observed outcomes, indicating the model's high potential for clinical translation (Figure 3D-F).

3.3. Multi-Resolution Assessment of the Immune Landscape

To explore the biological basis underlying the risk stratification, we analyzed the DEGs between the high- and low-risk groups. Functional enrichment analysis revealed clear associations with immune regulation. Specifically, we observed enrichment in Gene Ontology terms related to humoral immune response and antimicrobial humoral response. KEGG pathway analysis further highlighted signaling networks, including cytokine-cytokine receptor interaction, IL-17 signaling, and neuroactive ligand-receptor interaction (Figure 4A, B).

CIBERSORT and ssGSEA analyses also revealed marked differences in immune cell composition. The high-risk tumors displayed a distinct profile characterized by significantly lower infiltration of B cells, CD8⁺ T cells, and CD4⁺ T cells. In contrast, these high-risk samples were enriched with M2 macrophages (Figures 5A and 5B). This immunosuppressive pattern was corroborated by ESTIMATE analysis, where the low-risk group exhibited higher Immune and ESTIMATE scores and correspondingly lower tumor purity (Figure 5C-E), suggesting a more immunologically active environment.

3.4. Single-Cell Level Validation

We sought to validate these findings with greater precision by applying the Scissor algorithm to the

GSE161529 single-cell dataset. We successfully identified 666 cells associated with the high-risk phenotype (Scissor+) and 668 cells associated with the low-risk phenotype (Scissor-).

The cellular distribution mirrored our bulk analysis. The low-risk population was rich in T cells, B cells, endothelial cells, and fibroblasts, whereas the high-risk population was dominated by epithelial cells and macrophages (Figure 6A-C). Subclustering of T cells and macrophages revealed more specific information. The low-risk cells contained higher proportions of effector memory CD8+ T cells, central memory CD4+ T cells, and mast cells. Conversely, the high-risk cells were specifically enriched for regulatory T cells (Tregs) and M2 macrophages (Figure 6D-F). These single-cell findings align perfectly with the bulk transcriptomic landscape, reinforcing the biological relevance of the URS classification.

3.5. Genomic Alterations and Therapeutic Sensitivity

Genomic analysis revealed that the high-risk group was characterized by a higher Tumor Mutational Burden (TMB) (median 0.80 vs. 0.54 mutations/Mb; $p < 0.0001$) (Figure 7C-E) and a notably higher frequency of TP53 mutations (52% vs. 23%), suggesting a more unstable and aggressive tumor phenotype (Figure 7A, B).

Pharmacogenomic analyses indicated that the low-risk group was more responsive to drugs, even though high-risk tumors exhibited more aggressive features. Patients with low URS were predicted to be significantly more sensitive (lower IC50) to (Figure 8A-I):

1. CDK4/6 Inhibitors: Ribociclib, Palbociclib.
2. PARP Inhibitors: Olaparib, Talazoparib.
3. PI3K/AKT Pathway Inhibitors: Ipatasertib, Alpelisib.

4. Standard Chemotherapeutics: Epirubicin, Vincristine, Vinblastine.

These results suggest that the URS could serve as a companion diagnostic to select patients most likely to respond to these targeted and cytotoxic therapies.

3.6. Identification and Experimental Validation of SORBS1

To pinpoint the hub gene within our signature, we integrated Weighted Gene Co-expression Network Analysis (WGCNA) with machine learning algorithms, such as Random Forest and XGBoost. WGCNA identified the black module as the most strongly correlated with the URS ($r = 0.49$, $p = 5 \times 10^{-37}$) (Figure 9A-C). By intersecting the genes from this module with our six-gene signature, we highlighted SORBS1 and DTX1 as potential key candidates. Among these, SORBS1 consistently emerged as the top-ranked hub gene most strongly correlated with the risk phenotype (Figure 9D, E).

We validated the clinical relevance of SORBS1 at the protein level using immunohistochemistry (IHC) on an independent tissue microarray containing 136 breast tumors and 46 normal tissues. Typical IHC images are shown in Figure 10A. SORBS1 protein expression was significantly downregulated in tumor tissues (44.9% positivity) compared to normal breast tissues (71.7% positivity; $p = 0.0025$) (Figure 10B, Table 2). Importantly, survival analysis of the IHC cohort confirmed that patients with SORBS1-positive tumors experienced significantly longer overall survival ($p = 0.032$). This experimental evidence firmly establishes SORBS1 as a potential tumor suppressor and a protective prognostic biomarker in breast cancer.

4. Discussion

In this study, we established a six-gene ubiquitin-related prognostic signature that effectively stratified patients with breast cancer into high- and low-risk groups. Using the TCGA-BRCA cohort, the URS

demonstrated robust prognostic performance and remained an independent predictor of OS when combined with conventional clinicopathological variables. Beyond survival prediction, the URS was closely associated with distinct immunogenomic features, providing insight into how ubiquitin dysregulation may shape tumor aggressiveness and therapeutic responsiveness in breast cancer.

Comparison with Existing Prognostic Models Gene expression-based prognostic models for breast cancer have been widely reported; however, many are limited by large gene panels, lack of biological coherence, or insufficient validation (Györfy et al. 2015; Manjang et al. 2021). In contrast, the present study focuses specifically on ubiquitin-related genes, a functionally interconnected pathway with established relevance to cancer biology (Spano and Catara 2023b). The resulting signature comprises only six genes, enhancing its potential clinical feasibility. Importantly, we extended our analysis beyond bulk transcriptomic studies by integrating single-cell RNA sequencing data and validating the hub gene at the protein level using immunohistochemistry (Huang et al. 2023). This multi-layered strategy strengthens the biological credibility of the model and reduces the likelihood that the observed associations are driven solely by computational artifacts.

The Ubiquitin-Immune Connection: Unveiling the "Cold" Tumor Phenotype A notable finding of this work is the consistent association between high URS and an immunosuppressive tumor microenvironment. Patients in the high-risk group exhibited markedly reduced infiltration of CD8⁺ T cells and CD4⁺ T cells, together with an enrichment of M2 macrophages, indicating a shift toward immune exclusion (Yang et al. 2025a). These patterns were observed across multiple analytical frameworks, including CIBERSORT, ssGSEA, and ESTIMATE, and were further confirmed at single-cell resolution using the Scissor algorithm (Gan et al. 2024). Given the established role of M2 macrophages in promoting angiogenesis (Xu et al. 2025), extracellular matrix remodeling (Yang et al.

2025b), and immune evasion (Xia et al. 2025), their enrichment provides a plausible biological explanation for the poor prognosis observed in high-risk tumors (Jääskeläinen et al. 2024; Sami and Raza 2024). It is conceivable that dysregulated ubiquitination contributes to the degradation or functional suppression of immune-stimulatory molecules, such as antigen-presenting complexes (Lin et al. 2025) or cytokine signaling components (Hong et al. 2024), thereby facilitating immune escape (Jääskeläinen et al. 2024).

Therapeutic Implications: Towards Precision Medicine The integration of genomic and pharmacogenomic analyses revealed potential therapeutic implications of the URS. Although high-risk tumors displayed higher tumor mutational burden and frequent TP53 mutations—features often associated with genomic instability—they were predicted to be less responsive to several standard therapeutic agents (Kudo et al. 2024). In contrast, low-risk tumors demonstrated increased predicted sensitivity to CDK4/6 inhibitors (Kudo et al. 2024), PARP inhibitors (Zattarin et al. 2025), and anthracyclines (Chen et al. 2025). These findings suggest that the URS may serve as a useful companion biomarker for treatment stratification, helping identify patients more likely to benefit from targeted or cytotoxic therapies. For patients with high URS, alternative therapeutic strategies, such as rational combination regimens or immunomodulatory approaches, may be required to overcome intrinsic resistance.

SORBS1: A Promising Tumor Suppressor Through an integrative approach combining WGCNA and machine learning algorithms, SORBS1 emerged as a hub gene within the prognostic network. SORBS1 encodes a scaffold protein involved in cytoskeletal organization and cell adhesion, processes that are critical for maintaining epithelial integrity. Our immunohistochemical validation confirmed that SORBS1 protein expression is significantly reduced in breast cancer tissues compared with normal

controls, and that its preservation is associated with improved survival. These observations suggest that loss of SORBS1 may contribute to tumor progression by facilitating cellular detachment and invasive behavior. Although the precise ubiquitin-mediated mechanisms regulating SORBS1 remain to be elucidated, its consistent association with favorable outcomes supports its role as a protective biomarker.

Limitations Several limitations of this study should be acknowledged. First, the prognostic signature was derived from retrospective public datasets, and prospective validation in independent, multicenter cohorts is required before clinical implementation. Second, although single-cell RNA sequencing data were incorporated to refine the immune landscape, the available dataset was limited in size, potentially limiting the detection of rare cell populations. Finally, while protein-level validation was performed for the hub gene, functional experiments are needed to clarify the mechanistic roles of the remaining signature genes in ubiquitin-mediated immune regulation.

5. Conclusions

In conclusion, this study establishes a robust, multi-omics prognostic framework that goes beyond traditional clinical staging by linking ubiquitin-mediated processes to the tumor immune microenvironment. Our six-gene signature not only effectively stratifies patient survival but also reveals a critical association between the high ubiquitin-related risk group and an immunosuppressive, M2-macrophage-enriched phenotype. Furthermore, the identification of distinct therapeutic vulnerabilities—specifically associated with sensitivity to CDK4/6 and PARP inhibitors in the low-risk group — suggests immediate translational potential for guiding precision oncology. Finally, experimental validation of SORBS1 as a protective biomarker highlights its potential as a therapeutic

target. This finding offers a promising direction for future mechanistic studies. Collectively, these findings provide a valuable, clinically relevant tool for risk assessment and therapeutic decision-making in breast cancer.

Declarations

Ethics Approval and Consent

This study was conducted in accordance with the principles of the Declaration of Helsinki. The Institutional Review Board of the Second Affiliated Hospital of Harbin Medical University approved it. Informed consent was obtained from all subjects involved in the study.

Data Availability Statement

The datasets analyzed during the current study are available in public repositories. The TCGA-BRCA transcriptomic and clinicopathological data were obtained from the UCSC Xena GDC Hub ([https://xenabrowser.net/datapages/?cohort=GDC%20TCGA%20Breast%20Cancer%20\(BRCA\)](https://xenabrowser.net/datapages/?cohort=GDC%20TCGA%20Breast%20Cancer%20(BRCA))). The single cell RNA sequencing dataset GSE161529 is available in the Gene Expression Omnibus (GEO) repository at <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE161529>. The ubiquitin related gene set was retrieved from the iUUCD 2.0 database (<http://iuucd.biocuckoo.org/>).

Author Contributions

ZBF, YYZ, and SG contributed to the conception and design of the study. JQL, ML, YSL, WLC, YQD, ABH, MCL, and YLL collected the data. ZBF, FM, JRZ, YHS, JWL, and YHJ performed the data analysis. ZBF, JQL, and ML collectively conceptualized the manuscript. YYZ and SG edited the manuscript and provided critical comments. All authors contributed to manuscript writing and approved the final version of the manuscript.

Funding

This research was supported by the National Natural Science Foundation of China (82573913), the Natural Science Foundation of Heilongjiang Province (The Outstanding Youth Foundation) (YQ2024H016).

Conflict of Interest

The authors declare that they have no conflicts of interest.

Consent for Publication

Not applicable.

Clinical trial number: not applicable.

Abbreviations

AOD: Average Optical Density

AUC: Area under the curve

BC: Breast cancer

CIBERSORT: Cell-type Identification By Estimating Relative Subsets Of RNA Transcripts

CI: Confidence interval

COX: Cox proportional hazards regression

DCAF10 / DCAF13: DDB1- and CUL4-associated factors 10 / 13

DEGs: Differentially expressed genes

DTX1: Deltex E3 ubiquitin ligase 1

ER: Estrogen receptor

ESTIMATE: Estimation of Stromal and Immune Cells in Malignant Tumor Tissues Using Expression

Data

GDSC: Genomics of Drug Sensitivity in Cancer

GEO: Gene Expression Omnibus

GO: Gene Ontology

HER2: Human epidermal growth factor receptor 2

HPA: Human Protein Atlas

HR: Hazard ratio

IC50: Half-maximal inhibitory concentration

IHC: Immunohistochemistry

KEGG: Kyoto Encyclopedia of Genes and Genomes

LASSO: Least absolute shrinkage and selection operator

LONRF3: LON peptidase N-terminal domain and ring finger protein 3

OS: Overall survival

PTM: Post-translational modification

ROC: Receiver operating characteristic

scRNA-seq: Single-cell RNA sequencing

Scissor: Single-cell Integration tool identifying phenotype-associated cells

SORBS1: Sorbin and SH3 domain-containing protein 1

ssGSEA: Single-sample gene set enrichment analysis

TCGA: The Cancer Genome Atlas

TIME: Tumor immune microenvironment

TMB: Tumor mutational burden

TNBC: Triple-negative breast cancer

URS: Ubiquitin-related risk score

Ub-DEGs: Ubiquitin-related differentially expressed genes

UbRGs: Ubiquitin-related genes

UMAP: Uniform manifold approximation and projection

WGCNA: Weighted gene co-expression network analysis

XGBoost: Extreme gradient boosting machine-learning algorithm

References

- Bai J, Gao Y, Zhang G (2025) The treatment of breast cancer in the era of precision medicine. *Cancer Biol Med* 1–26. <https://doi.org/10.20892/j.issn.2095-3941.2024.0510>
- Chen N, Freeman JQ, Yarlagadda S, Atmakuri A, Kalinsky K, Pusztai L, Sparano JA, Huo D, Nanda R, Howard FM (2025) Impact of anthracyclines in genomic high-risk, node-negative, HR-positive/HER2-negative breast cancer. *Ann Oncol* 36:1356–1365. <https://doi.org/10.1016/j.annonc.2025.08.002>
- Dhatchinamoorthy K, Colbert JD, Rock KL (2021) Cancer Immune Evasion Through Loss of MHC Class I Antigen Presentation. *Front Immunol* 12:636568. <https://doi.org/10.3389/fimmu.2021.636568>
- Diamandis EP (2012) The failure of protein cancer biomarkers to reach the clinic: why, and what can be done to address the problem? *BMC Med* 10:87. <https://doi.org/10.1186/1741-7015-10-87>
- Feng K, Di Y, Han M, Yan W, Wang Y (2024) SORBS1 inhibits epithelial to mesenchymal transition (EMT) of breast cancer cells by regulating PI3K/AKT signaling and macrophage phenotypic polarization. *Aging*. <https://doi.org/10.18632/aging.205632>
- Gan D, Zhu Y, Lu X, Li J (2024) SCIPAC: quantitative estimation of cell-phenotype associations. *Genome Biol* 25:119. <https://doi.org/10.1186/s13059-024-03263-1>
- Gao T, Sun Y, Leng P, Liu D, Guo Q, Li J (2025) CDK4/6 inhibitors in breast cancer therapy: mechanisms of drug resistance and strategies for treatment. *Front Pharmacol* 16:1549520. <https://doi.org/10.3389/fphar.2025.1549520>
- Györfy B, Hatzis C, Sanft T, Hofstatter E, Aktas B, Pusztai L (2015) Multigene prognostic tests in breast cancer: past, present, future. *Breast Cancer Res* 17:11.

<https://doi.org/10.1186/s13058-015-0514-2>

- Hao Y, Hao S, Andersen-Nissen E, Mauck WM, Zheng S, Butler A, Lee MJ, Wilk AJ, Darby C, Zager M, Hoffman P, Stoeckius M, Papalexi E, Mimitou EP, Jain J, Srivastava A, Stuart T, Fleming LM, Yeung B, Rogers AJ, McElrath JM, Blish CA, Gottardo R, Smibert P, Satija R (2021) Integrated analysis of multimodal single-cell data. *Cell* 184:3573–3587.e29. <https://doi.org/10.1016/j.cell.2021.04.048>
- Hong Z, Liu F, Zhang Z (2024) Ubiquitin modification in the regulation of tumor immunotherapy resistance mechanisms and potential therapeutic targets. *Exp Hematol Oncol* 13:91. <https://doi.org/10.1186/s40164-024-00552-0>
- Hu C, Li T, Xu Y, Zhang X, Li F, Bai J, Chen J, Jiang W, Yang K, Ou Q, Li X, Wang P, Zhang Y (2023) CellMarker 2.0: an updated database of manually curated cell markers in human/mouse and web tools based on scRNA-seq data. *Nucleic Acids Res* 51:D870–D876. <https://doi.org/10.1093/nar/gkac947>
- Huang L, Qin S, Xia L, Ma F, Chen L (2023) Integrative single-cell sequencing analysis distinguishes survival-associated cells from the breast cancer microenvironment. *Cancer Med* 12:12896–12911. <https://doi.org/10.1002/cam4.5892>
- Huang Y, Gao Y, Lin Z, Miao H (2025) Involvement of the ubiquitin-proteasome system in the regulation of the tumor microenvironment and progression. *Genes Dis* 12:101240. <https://doi.org/10.1016/j.gendis.2024.101240>
- Jääskeläinen MM, Tumelius R, Hämäläinen K, Rilla K, Oikari S, Rönkä A, Selander T, Mannermaa A, Tiainen S, Auvinen P (2024) High Numbers of CD163+ Tumor-Associated Macrophages Predict Poor Prognosis in HER2+ Breast Cancer. *Cancers* 16:634. <https://doi.org/10.3390/cancers16030634>
- Jin X, Qiu X, Huang Y, Zhang H, Chen K (2022) miR-223-3p carried by cancer-associated fibroblast microvesicles targets SORBS1 to modulate the progression of gastric cancer. *Cancer Cell Int* 22:96. <https://doi.org/10.1186/s12935-022-02513-1>
- Kim J, Harper A, McCormack V, Sung H, Houssami N, Morgan E, Mutebi M, Garvey G, Soerjomataram I, Fidler-Benaoudia MM (2025) Global patterns and trends in breast cancer incidence and mortality across 185 countries. *Nat Med* 31:1154–1162. <https://doi.org/10.1038/s41591-025-03502-3>
- Koo S-Y, Park E-J, Noh H-J, Jo S-M, Ko B-K, Shin H-J, Lee C-W (2023) Ubiquitination Links DNA Damage and Repair Signaling to Cancer Metabolism. *Int J Mol Sci* 24:8441. <https://doi.org/10.3390/ijms24098441>
- Kudo R, Safonov A, Jones C, Moiso E, Dry JR, Shao H, Nag S, Da Silva EM, Yildirim SY, Li Q, O’Connell E, Patel P, Will M, Fushimi A, Benitez M, Bradic M, Fan L, Nakshatri H, Sudhan DR, Denz CR, Huerga Sanchez I, Reis-Filho JS, Goel S, Koff A, Weigelt B, Khan QJ, Razavi P, Chandarlapaty S (2024) Long-term breast cancer response to CDK4/6 inhibition defined by

TP53-mediated geroconversion. *Cancer Cell* 42:1919-1935.e9.
<https://doi.org/10.1016/j.ccell.2024.09.009>

Langfelder P, Horvath S (2008) WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 9:559. <https://doi.org/10.1186/1471-2105-9-559>

Lemma EY, Letian A, Altorki NK, McGraw TE (2023) Regulation of PD-L1 Trafficking from Synthesis to Degradation. *Cancer Immunol Res* 11:866–874.
<https://doi.org/10.1158/2326-6066.CIR-22-0953>

Li J, Jiang Y, Ma M, Wang L, Jing M, Yang Z, Zhang M, Chen K, Fan J (2025) Epithelial cell diversity and immune remodeling in bladder cancer progression: insights from single-cell transcriptomics. *J Transl Med* 23:135. <https://doi.org/10.1186/s12967-025-06138-6>

Lin P, Lin Y, Chen X, Zhao X, Cui L (2025) Decoding MHC loss: Molecular mechanisms and implications for immune resistance in cancer. *Clin Transl Med* 15:e70403.
<https://doi.org/10.1002/ctm2.70403>

Ma H-Y, Cao J-M, Zhang Y-Y, Yang J-S, Wang X, Yu Y (2025) The Clinical Prediction Value of the Ubiquitination Model Reflecting the Microenvironment Infiltration and Drug Sensitivity in Breast Cancer. *J Cancer* 16:784–801. <https://doi.org/10.7150/jca.101525>

Maeser D, Gruener RF, Huang RS (2021) oncoPredict: an R package for predicting in vivo or cancer patient drug response and biomarkers from cell line screening data. *Brief Bioinform* 22:bbab260. <https://doi.org/10.1093/bib/bbab260>

Manjang K, Tripathi S, Yli-Harja O, Dehmer M, Glazko G, Emmert-Streib F (2021) Prognostic gene expression signatures of breast cancer are lacking a sensible biological meaning. *Sci Rep* 11:156. <https://doi.org/10.1038/s41598-020-79375-y>

Mayakonda A, Lin D-C, Assenov Y, Plass C, Koeffler HP (2018) Maftools: efficient and comprehensive analysis of somatic variants in cancer. *Genome Res* 28:1747–1756.
<https://doi.org/10.1101/gr.239244.118>

Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, Hoang CD, Diehn M, Alizadeh AA (2015) Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods* 12:453–457. <https://doi.org/10.1038/nmeth.3337>

Robinson MD, McCarthy DJ, Smyth GK (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26:139–140.
<https://doi.org/10.1093/bioinformatics/btp616>

Sabit H, Abbas S, El-Safoury MT, Madkour EM, Mahmoud S, Abdel-Ghany S, Albrahim Y, Al-Dhuayan IS, Rashwan S, El-Hashash A, Arneth B (2025) Antibody–Drug Conjugates in Breast Cancer: Navigating Innovations, Overcoming Resistance, and Shaping Future Therapies. *Biomedicines* 13:2227. <https://doi.org/10.3390/biomedicines13092227>

- Saha S, Mahapatra S, Khanra S, Mishra B, Swain B, Malhotra D, Saha S, Panda VK, Kumari K, Jena S, Thakur S, Singh PK, Kundu GC (2025) Decoding breast cancer treatment resistance through genetic, epigenetic, and immune-regulatory mechanisms: from molecular insights to translational perspectives. *Cancer Drug Resist.* <https://doi.org/10.20517/cdr.2025.69>
- Sami A, Raza A (2024) Reprogramming the tumor microenvironment – macrophages emerge as key players in breast cancer immunotherapy. *Front Immunol* 15:1457491. <https://doi.org/10.3389/fimmu.2024.1457491>
- Simon N, Friedman J, Hastie T, Tibshirani R (2011) Regularization Paths for Cox's Proportional Hazards Model via Coordinate Descent. *J Stat Softw* 39. <https://doi.org/10.18637/jss.v039.i05>
- Spano D, Catara G (2023a) Targeting the Ubiquitin–Proteasome System and Recent Advances in Cancer Therapy. *Cells* 13:29. <https://doi.org/10.3390/cells13010029>
- Spano D, Catara G (2023b) Targeting the Ubiquitin–Proteasome System and Recent Advances in Cancer Therapy. *Cells* 13:29. <https://doi.org/10.3390/cells13010029>
- Sun D, Guan X, Moran AE, Wu L-Y, Qian DZ, Schedin P, Dai M-S, Danilov AV, Alumkal JJ, Adey AC, Spellman PT, Xia Z (2022) Identifying phenotype-associated subpopulations by integrating bulk and single-cell sequencing data. *Nat Biotechnol* 40:527–538. <https://doi.org/10.1038/s41587-021-01091-3>
- Wright G, Tan B, Rosenwald A, Hurt EH, Wiestner A, Staudt LM (2003) A gene expression-based method to diagnose clinically distinct subgroups of diffuse large B cell lymphoma. *Proc Natl Acad Sci* 100:9991–9996. <https://doi.org/10.1073/pnas.1732008100>
- Wu SZ, Al-Eryani G, Roden DL, Junankar S, Harvey K, Andersson A, Thennavan A, Wang C, Torpy JR, Bartonicek N, Wang T, Larsson L, Kaczorowski D, Weisenfeld NI, Uytingco CR, Chew JG, Bent ZW, Chan C-L, Gnanasambandapillai V, Dutertre C-A, Gluch L, Hui MN, Beith J, Parker A, Robbins E, Segara D, Cooper C, Mak C, Chan B, Warriar S, Ginhoux F, Millar E, Powell JE, Williams SR, Liu XS, O'Toole S, Lim E, Lundeberg J, Perou CM, Swarbrick A (2021) A single-cell and spatially resolved atlas of human breast cancers. *Nat Genet* 53:1334–1347. <https://doi.org/10.1038/s41588-021-00911-1>
- Xia W, Zhang X, Wang Y, Huang Z, Guo X, Fang L (2025) Progress in targeting tumor-associated macrophages in cancer immunotherapy. *Front Immunol* 16:1658795. <https://doi.org/10.3389/fimmu.2025.1658795>
- Xu J, Ding L, Mei J, Hu Y, Kong X, Dai S, Bu T, Xiao Q, Ding K (2025) Dual roles and therapeutic targeting of tumor-associated macrophages in tumor microenvironments. *Signal Transduct Target Ther* 10:268. <https://doi.org/10.1038/s41392-025-02325-5>
- Yang H, Sun T, Sun Z, Wang H, Liu D, Wu D, Qin T, Zhou M (2025a) Unravelling the role of ubiquitin-specific proteases in breast carcinoma: insights into tumour progression and immune microenvironment modulation. *World J Surg Oncol* 23:60. <https://doi.org/10.1186/s12957-025-03667-8>

- Yang Y, Li S, To KKW, Zhu S, Wang F, Fu L (2025b) Tumor-associated macrophages remodel the suppressive tumor immune microenvironment and targeted therapy for immunotherapy. *J Exp Clin Cancer Res* 44:145. <https://doi.org/10.1186/s13046-025-03377-9>
- Yoshihara K, Shahmoradgoli M, Martínez E, Vegesna R, Kim H, Torres-Garcia W, Treviño V, Shen H, Laird PW, Levine DA, Carter SL, Getz G, Stemke-Hale K, Mills GB, Verhaak RGW (2013) Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat Commun* 4:2612. <https://doi.org/10.1038/ncomms3612>
- Yu G, Wang L-G, Han Y, He Q-Y (2012) clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. *OMICS J Integr Biol* 16:284–287. <https://doi.org/10.1089/omi.2011.0118>
- Zattarin E, Marra A, Palazzo A, Griguolo G, Vernieri C, Etessami J, Pontolillo L, Landa G, Daneri A, De Monte M, Cuoghi Costantini R, Tenedini E, Ponzoni O, Razeti MG, Sposetti C, Barbieri E, Manni M, Caggia F, Cortesi L, Curigliano G, Bria E, Dominici M, Guarneri V, Lambertini M, Toss A (2025) Real-world effectiveness of PARP inhibitors after CDK4/6 inhibitor therapy in BRCA-mutated HR-positive/HER2-negative advanced breast cancer. *Npj Breast Cancer*. <https://doi.org/10.1038/s41523-025-00859-z>
- Zhou J, Xu Y, Lin S, Guo Y, Deng W, Zhang Y, Guo A, Xue Y (2018) iUUCD 2.0: an update with rich annotations for ubiquitin and ubiquitin-like conjugations. *Nucleic Acids Res* 46:D447–D453. <https://doi.org/10.1093/nar/gkx1041>
- Zhu S, Zhang X, Liu W, Zhou Z, Xiong S, Li J, Chen X, Peng C (2025) Ubiquitination in cancer: mechanisms and therapeutic opportunities. *Cancer Commun* 45:1128–1161. <https://doi.org/10.1002/cac2.70044>

Tables

Table 1. Baseline characteristics of BC patients in the Training set and the Testing set.

Variables	Training set(N=593)		Testing set(N=254)		P-value
	^a N	%	^a N	%	
Age					0.48
<50	177	29.8	69	27.2	
≥50	416	70.2	185	72.8	
T Stage					0.888

T1	165	27.8	64	25.2	
T2	342	57.7	151	59.4	
T3	66	11.1	30	11.8	
T4	20	3.4	9	3.5	
^b N Stage					0.195
N0	283	47.7	134	52.8	
N1	210	35.4	74	29.1	
N2	67	11.3	26	10.2	
N3	33	5.6	20	7.9	
M Stage					0.255
M0	585	98.7	247	97.2	
M1	8	1.3	7	2.8	
Stage					0.516
Stage I	110	18.5	45	17.7	
Stage II	342	57.7	149	58.7	
Stage III	133	22.4	53	20.9	
Stage IV	8	1.3	7	2.8	

BC: breast cancer; ^aN: number; %: percent; ^bN: lymph node.

Table 2. Comparison of SORBS1 expression between BC patients and normal tissues.

	BC (N=136)		Normal (N=46)		P-value
	N	%	N	%	

SORBS1					0.0025
Positive	61	44.9	33	71.7	
Negative	75	55.1	13	28.3	

BC: breast cancer; N: number; %: percent.

Figure Legends

Figure 1. Screening and Identification of Ubiquitin-Related Prognostic Genes.

(A) Volcano plot visualizing differentially expressed genes (DEGs) between tumor and normal tissues. (B) Venn diagram illustrating the intersection between identified DEGs and the ubiquitin-related gene (UbRG) set. (C) LASSO coefficient profiles for the candidate UbRGs. (D) Cross-validation plot identifying the optimal penalty parameter (lambda) for the LASSO model. (E) Forest plot showing UbRGs significantly associated with overall survival in univariate Cox regression. (F) Comparison of expression levels for the six signature genes between high- and low-risk groups. (G) Representative immunohistochemical staining images of the six UbRGs in normal and breast cancer tissues obtained from the Human Protein Atlas (HPA), a publicly available database.

Figure 2. Prognostic Performance of the Six-Gene Signature Across Cohorts.

(A–B) Distribution of risk scores and corresponding patient survival status in the training cohort. (C) Kaplan-Meier curve showing overall survival (OS) in the training cohort. (D) Time-dependent ROC curves assessing predictive accuracy at 1, 7, and 10 years in the training cohort. (E–F) Kaplan-Meier analysis and time-dependent ROC curves in the internal testing cohort. (G–H) Kaplan-Meier analysis and time-dependent ROC curves in the entire TCGA-BRCA cohort. Note: Patients were stratified into high-risk (red) and low-risk (blue/green) groups based on the median risk score.

Figure 3. Independence and Clinical Utility of the Risk Signature.

(A) Univariate and (B) multivariate Cox regression analyses evaluating the independence of the risk

score relative to standard clinicopathological variables. (C) A predictive nomogram integrating the six-gene risk score with patient age, T stage, and N stage to estimate 1-, 7-, and 10-year overall survival probabilities. (D–F) Calibration plots demonstrating the concordance between nomogram-predicted probabilities and observed survival rates at 1, 7, and 10 years.

Figure 4. Functional Enrichment Analysis of Differentially Expressed Genes.

Bar plots summarizing the functional enrichment results for genes differentially expressed between high- and low-risk groups, including (A) Gene Ontology (GO) terms and (B) Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways.

Figure 5. Immune Landscape Differences Between Risk Groups.

(A) Relative proportions of 22 immune cell phenotypes estimated by the CIBERSORT algorithm. (B) Enrichment scores for 28 immune cell populations calculated via ssGSEA. (C–E) Comparison of ESTIMATE scores, immune scores, and tumor purity between the high- and low-risk groups. ns, not significant; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

Figure 6. Single-Cell Resolution Analysis and Scissor-Identified Subpopulations (GSE161529).

(A) UMAP plot visualizing the major cell types identified in the single-cell dataset. (B) UMAP projection distinguishing Scissor+ (high-risk associated) and Scissor– (low-risk associated) cells. (C) Bar plot showing the cellular composition of Scissor+ versus Scissor– populations. (D) UMAP visualization of the sub-clustered immune compartment. (E) Distribution of Scissor+ and Scissor– cells within the immune sub-clusters. (F) Proportions of specific immune subsets within the Scissor+ and Scissor– groups. (G–H) UMAP plots highlighting distinct cell clusters and the expression of canonical marker genes. (I–J) Dot plots illustrating marker gene expression profiles in the sub-clustered immune cells.

Figure 7. Somatic Mutation Landscape and Tumor Mutational Burden.

(A–B) Waterfall plots depicting the top mutated genes in the (A) low-risk and (B) high-risk groups. (C–D) Distribution of Tumor Mutational Burden (TMB) in the low-risk and high-risk groups, respectively. (E) Statistical comparison of TMB levels between the high- and low-risk groups.

Figure 8. Predicted Chemotherapeutic Sensitivity.

Box plots comparing the estimated IC50 values for (A) alpelisib, (B) ribociclib, (C) olaparib, (D) vincristine, (E) ipatasertib, (F) epirubicin, (G) palbociclib, (H) talazoparib, and (I) vinblastine between high- and low-risk groups. Note: A lower IC50 value indicates higher predicted drug sensitivity. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

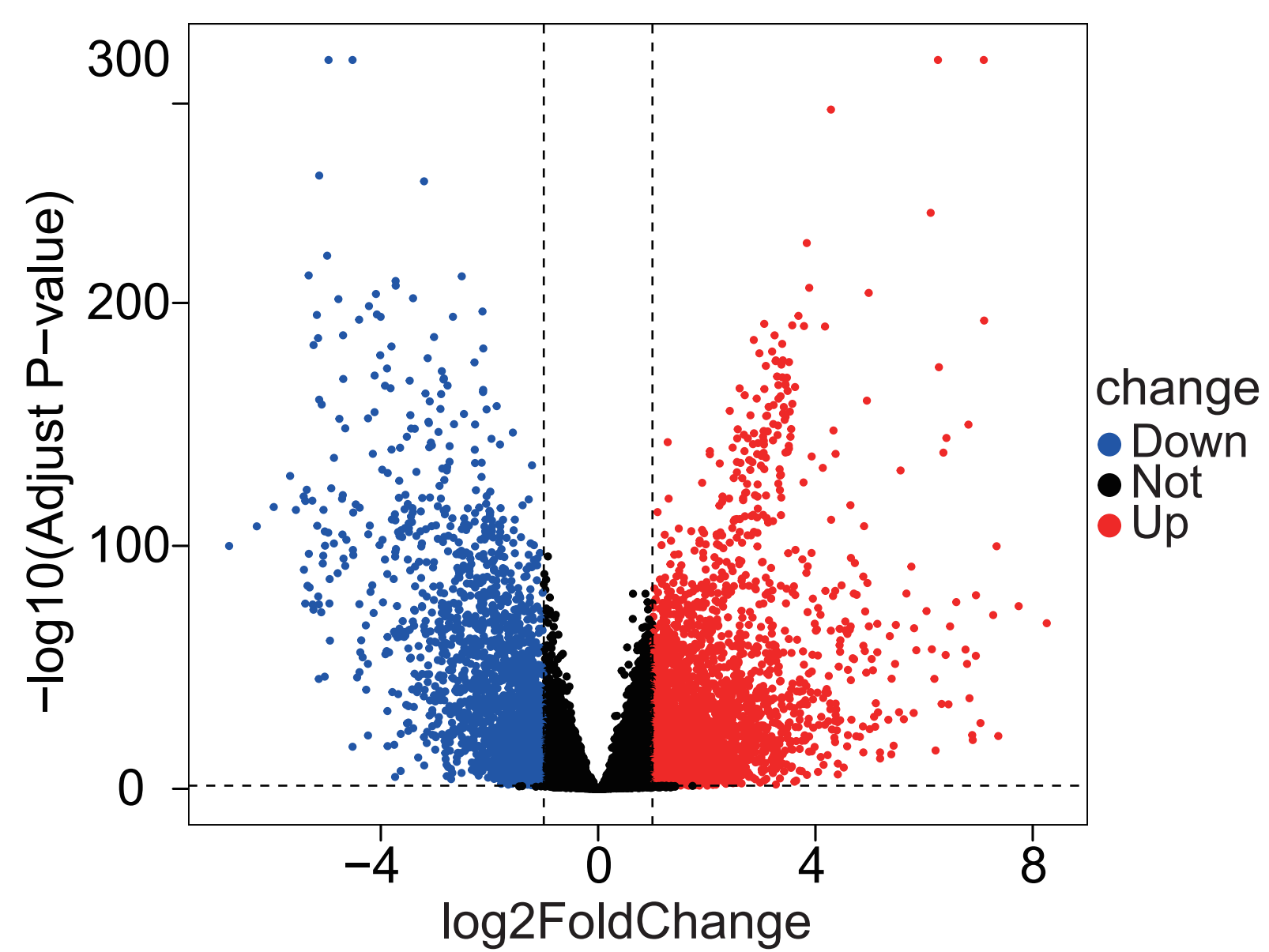
Figure 9. Identification of SORBS1 as the Hub Gene.

(A) Sample clustering dendrogram paired with a heatmap of risk scores. (B) Analysis of the scale-free fit index to determine the optimal soft-thresholding power ($\beta = 4$). (C) Module–trait relationship heatmap, highlighting the strong correlation between the black module and the risk score. (D) Feature importance ranking of the six signature genes derived from the Random Forest algorithm. (E) Feature importance ranking derived from the XGBoost algorithm.

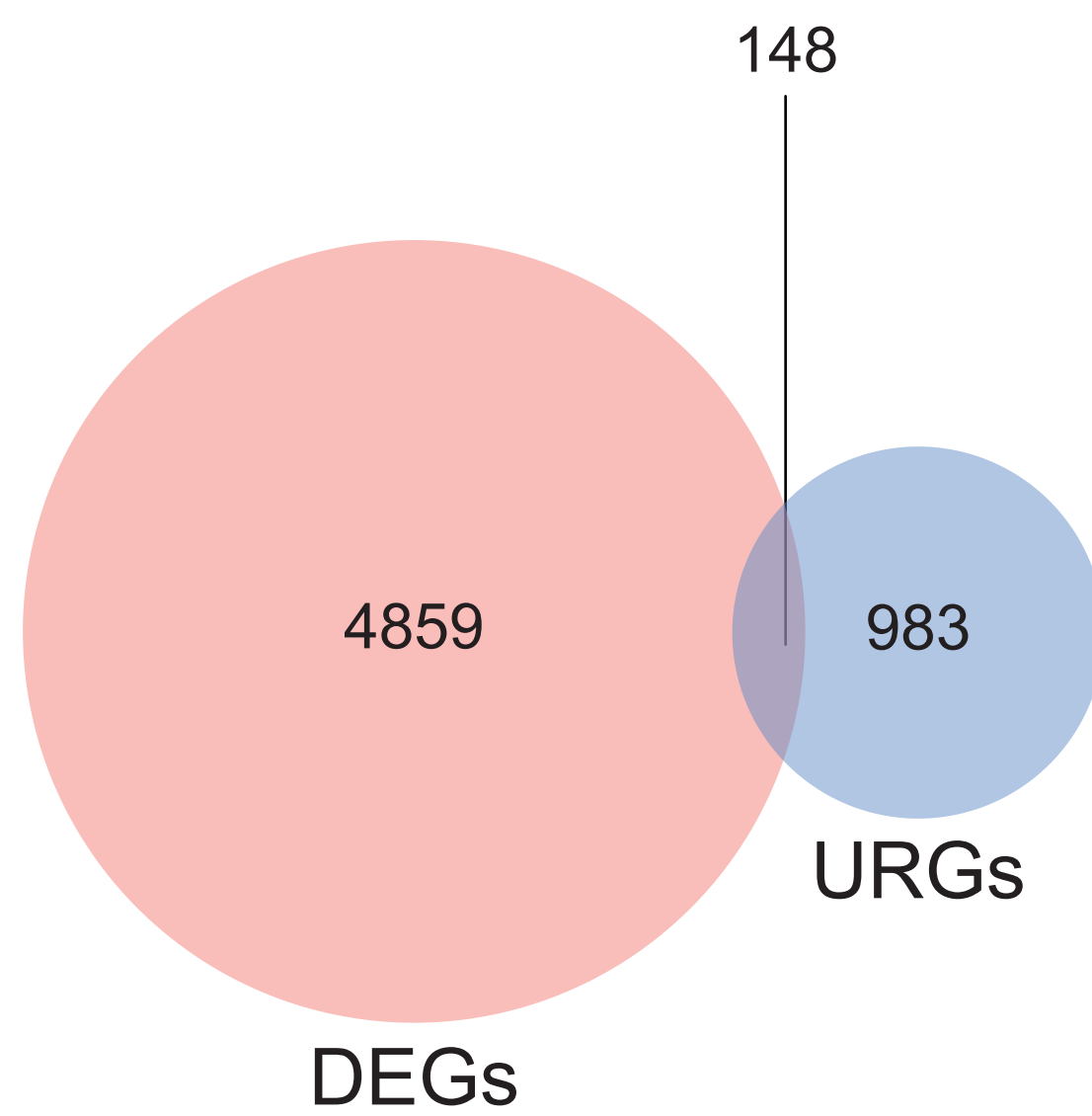
Figure 10. Immunohistochemical Validation of SORBS1.

(A) Representative immunohistochemical (IHC) staining images showing SORBS1 expression in normal breast tissue versus breast cancer tissue (magnification: 4 $\times$ and 40 $\times$). (B) Bar chart comparing the positivity rate of SORBS1 in normal ($n = 46$) and tumor ($n = 136$) tissues. (C) Kaplan–Meier curve analyzing the association between SORBS1 protein expression and overall survival in the validation cohort ($p = 0.032$).

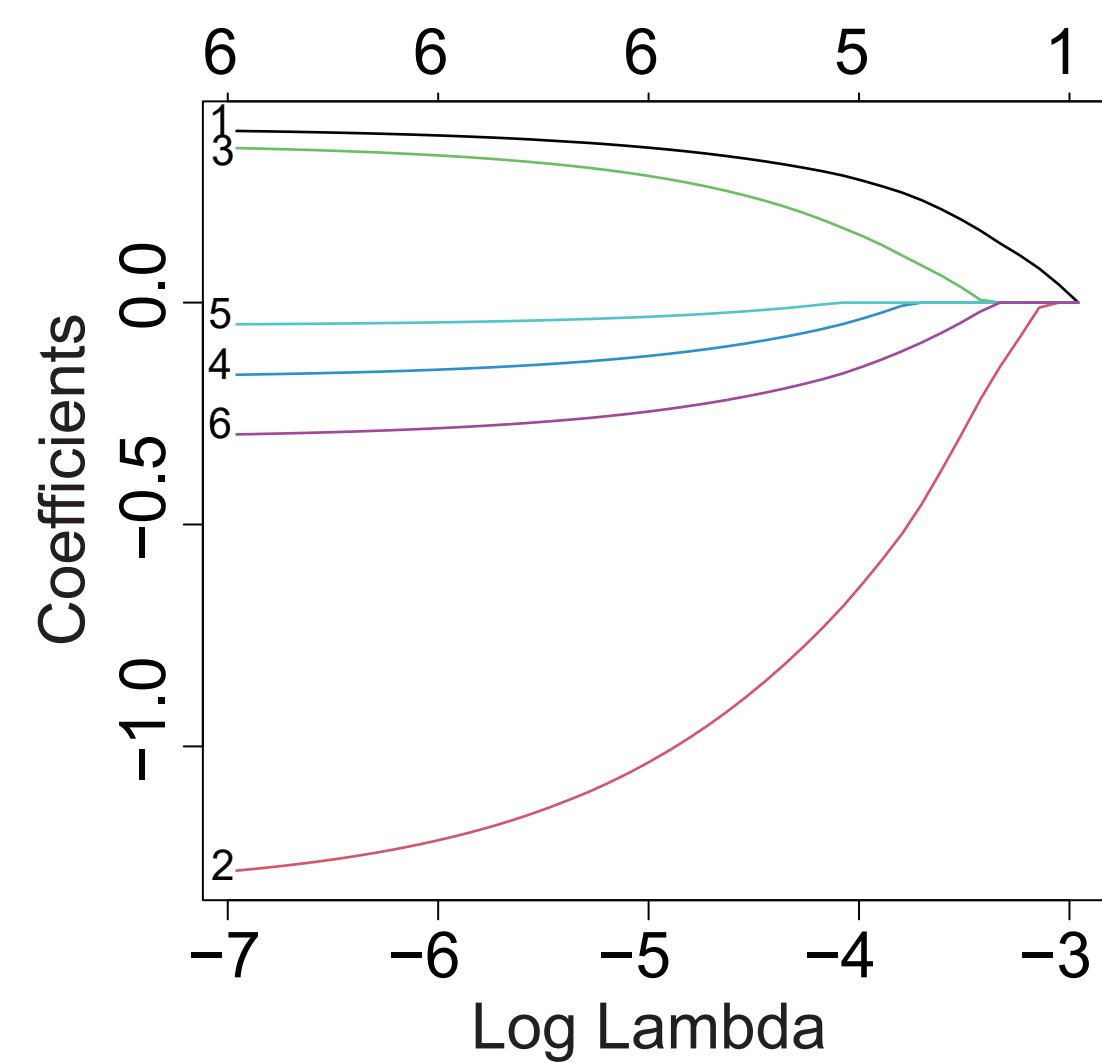
A



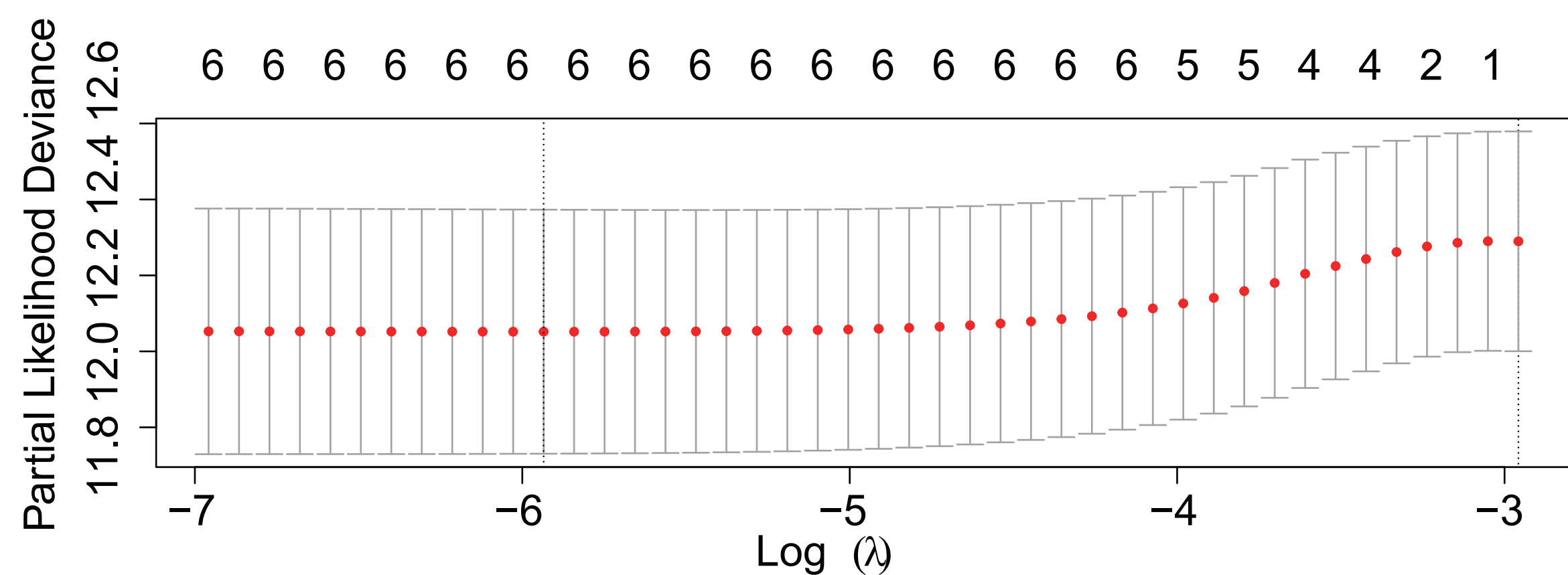
B



C



D

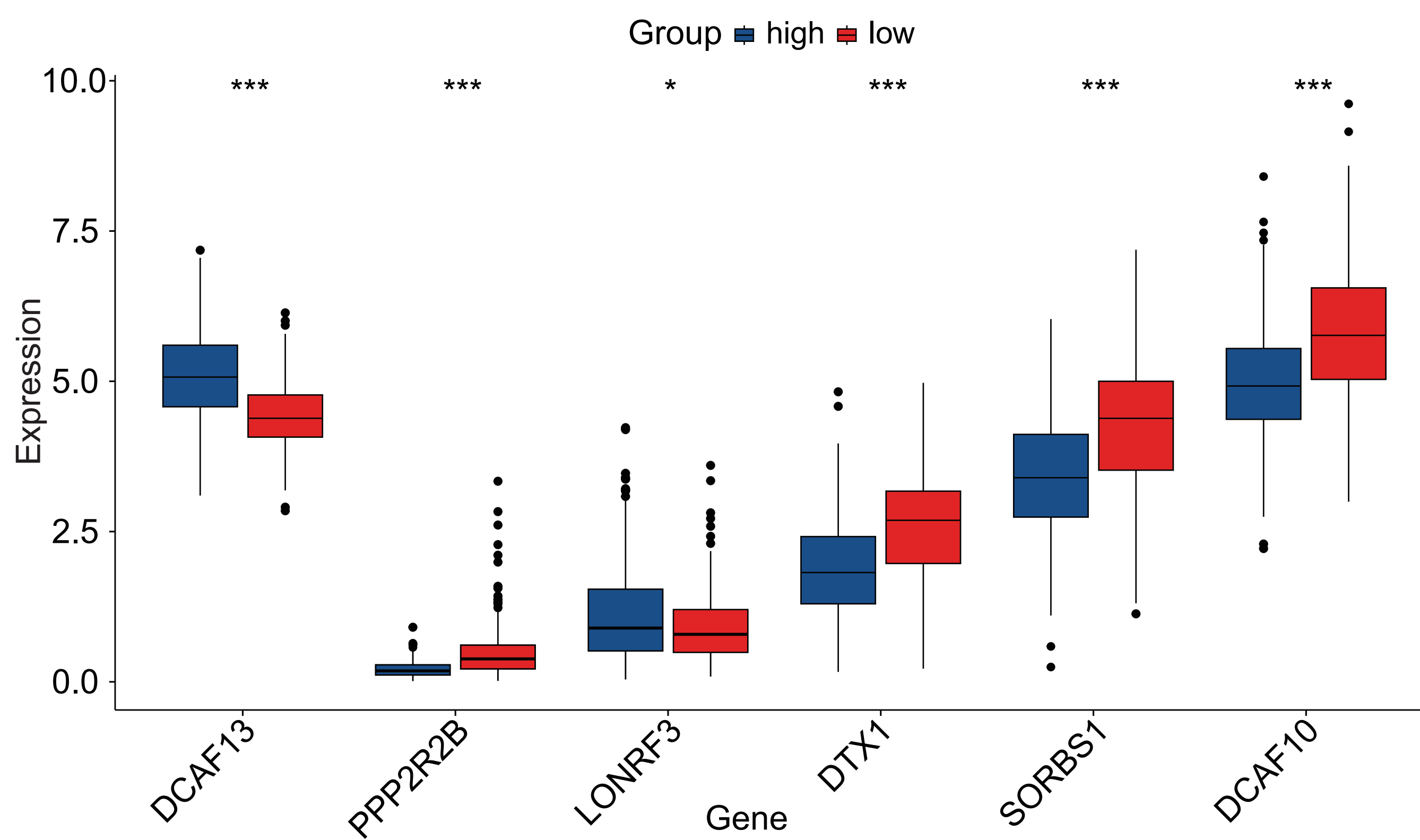


E

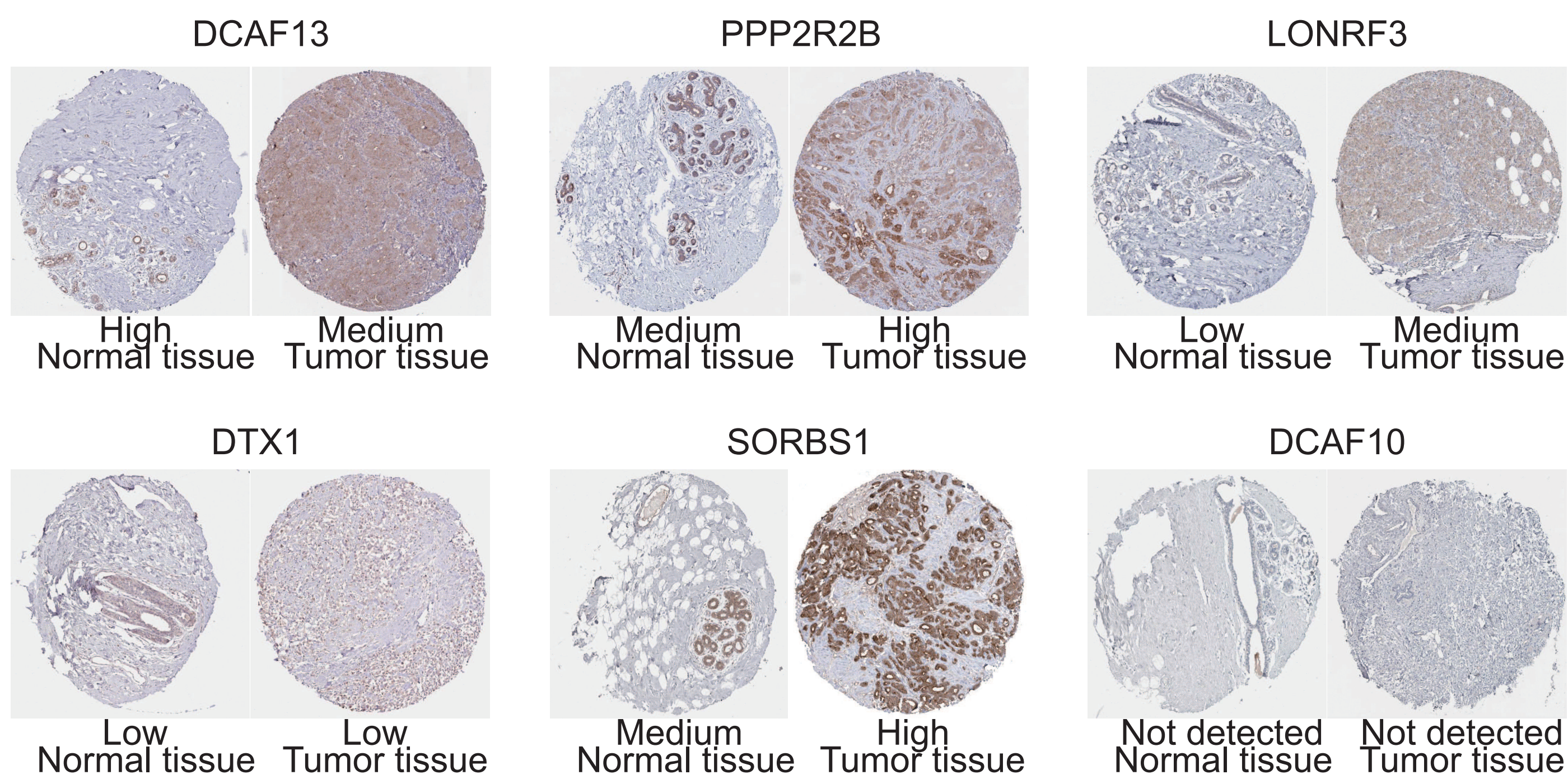
Gene	HR	lower 95%CI	upper 95%CI	pvalue
DCAF13	1.534	1.182	1.991	0.001
PPP2R2B	0.270	0.107	0.677	0.005
LONRF3	1.317	1.015	1.708	0.038
DTX1	0.783	0.619	0.991	0.042
SORBS1	0.822	0.678	0.995	0.044
DCAF10	0.816	0.666	0.999	0.049

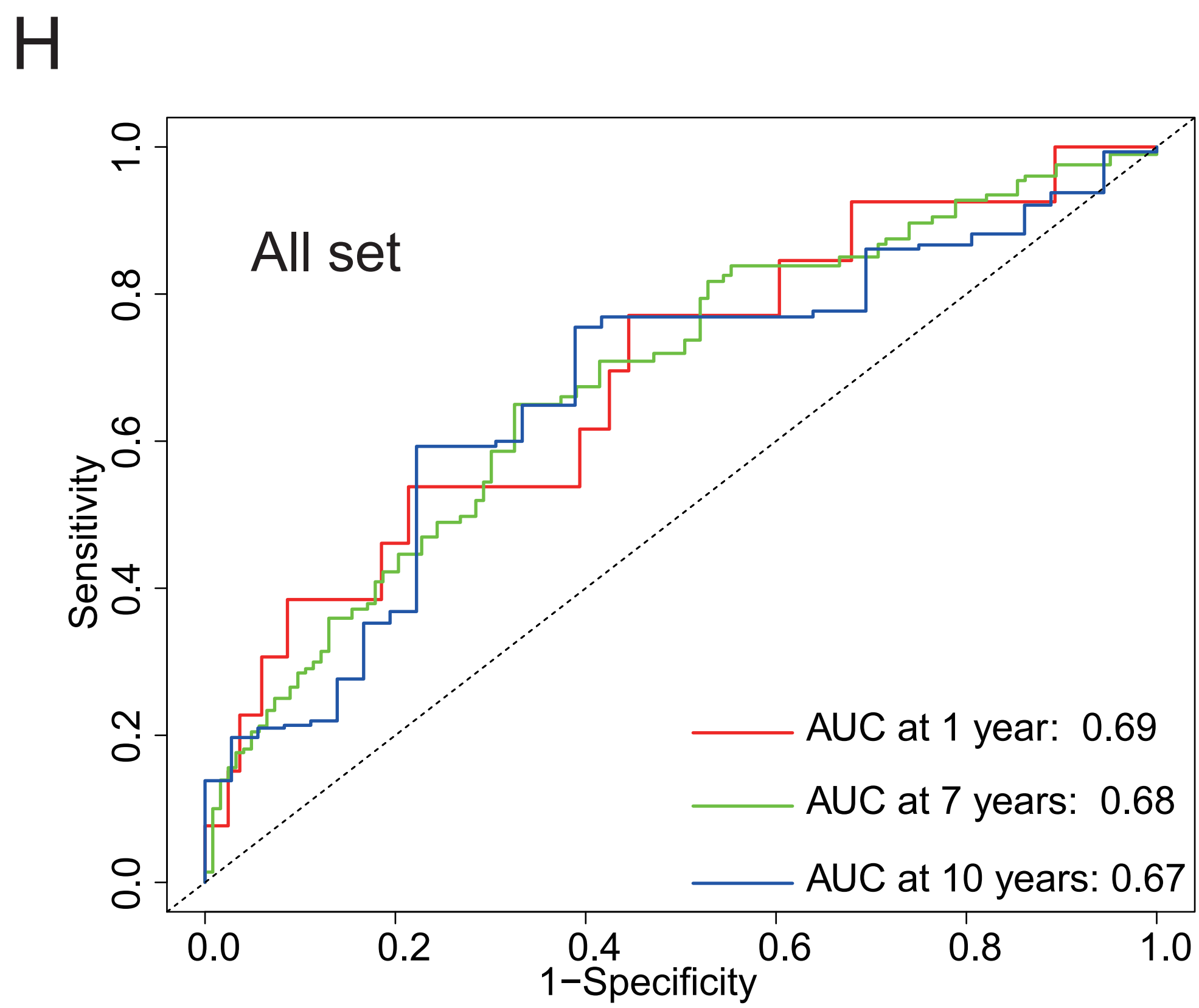
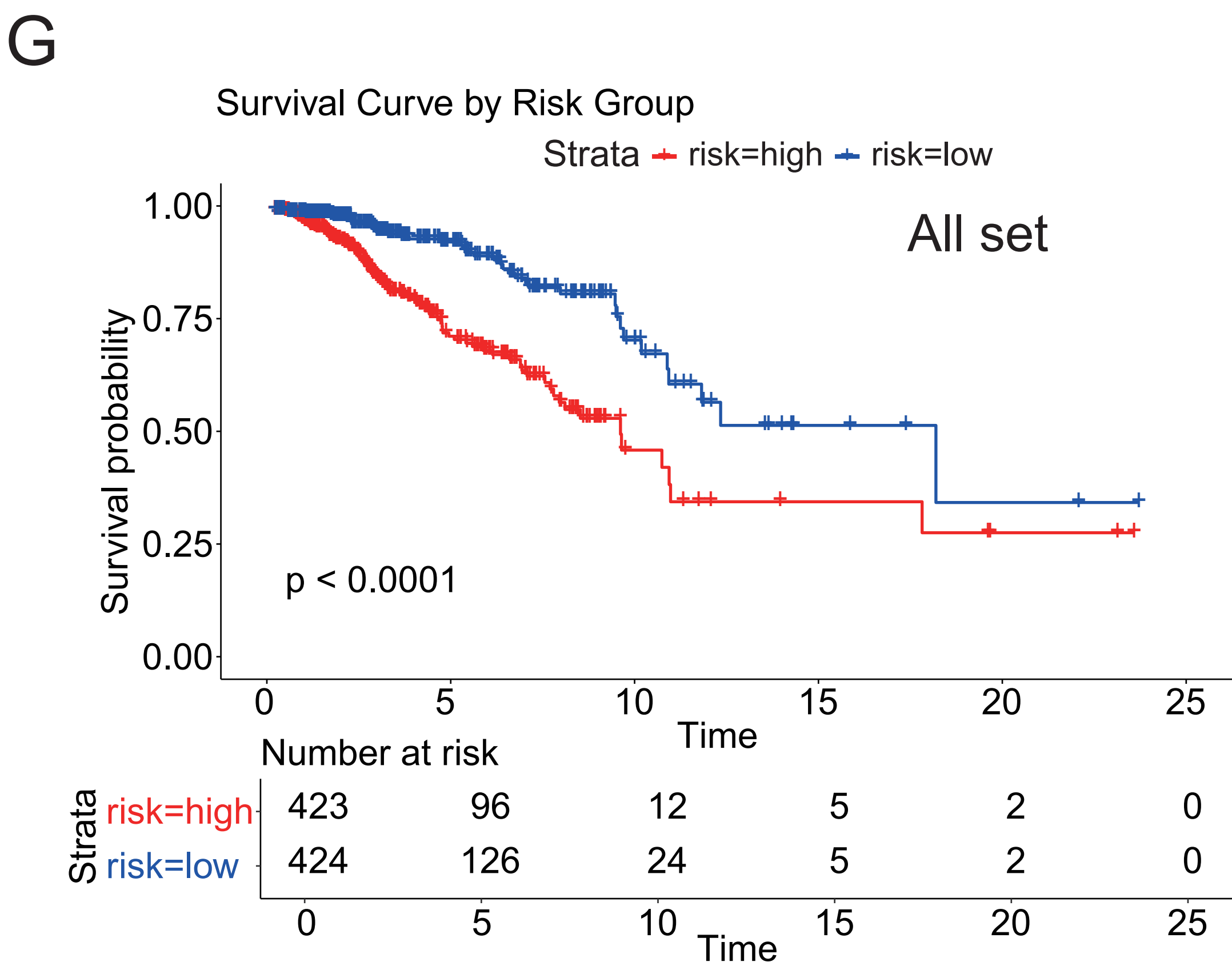
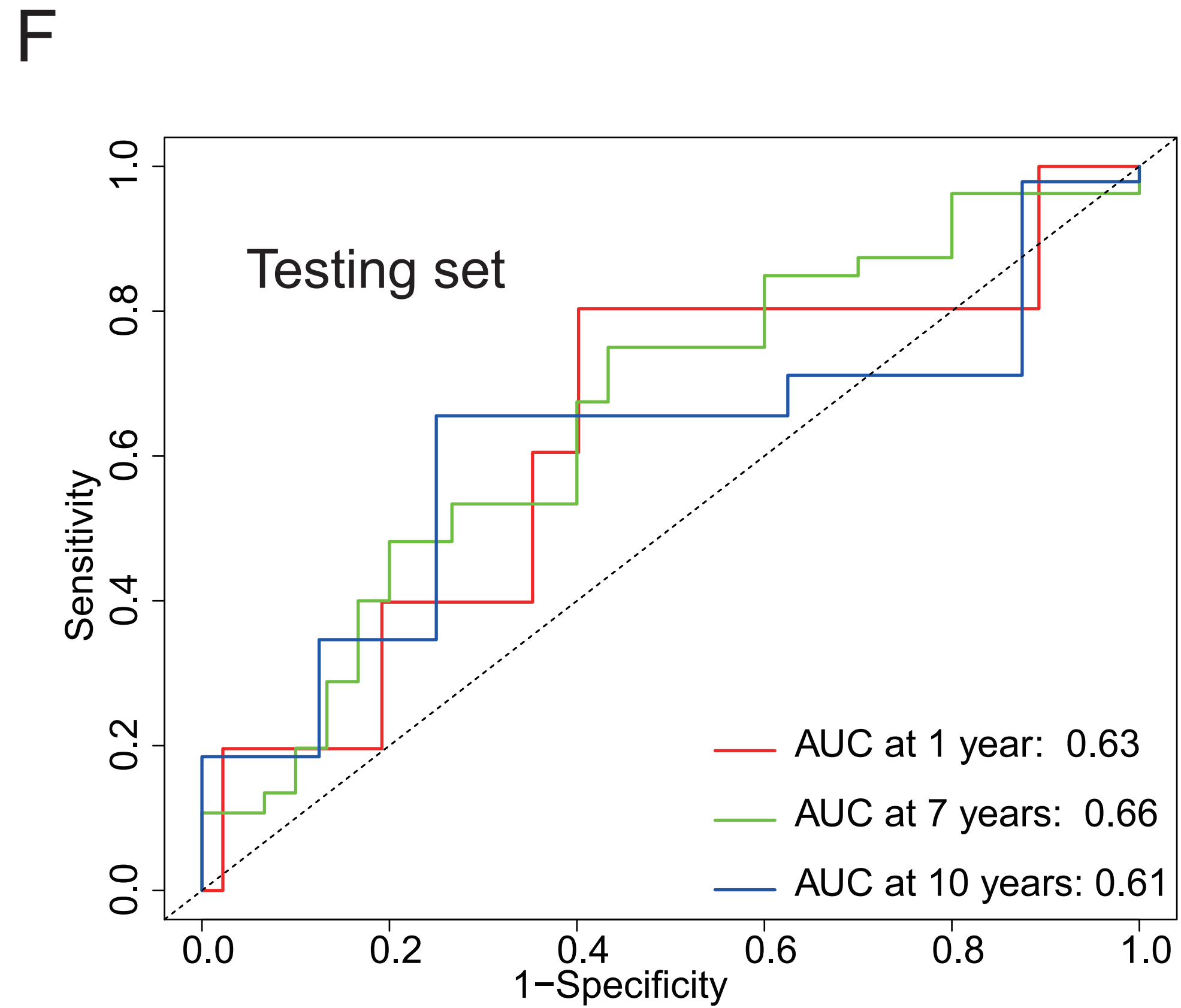
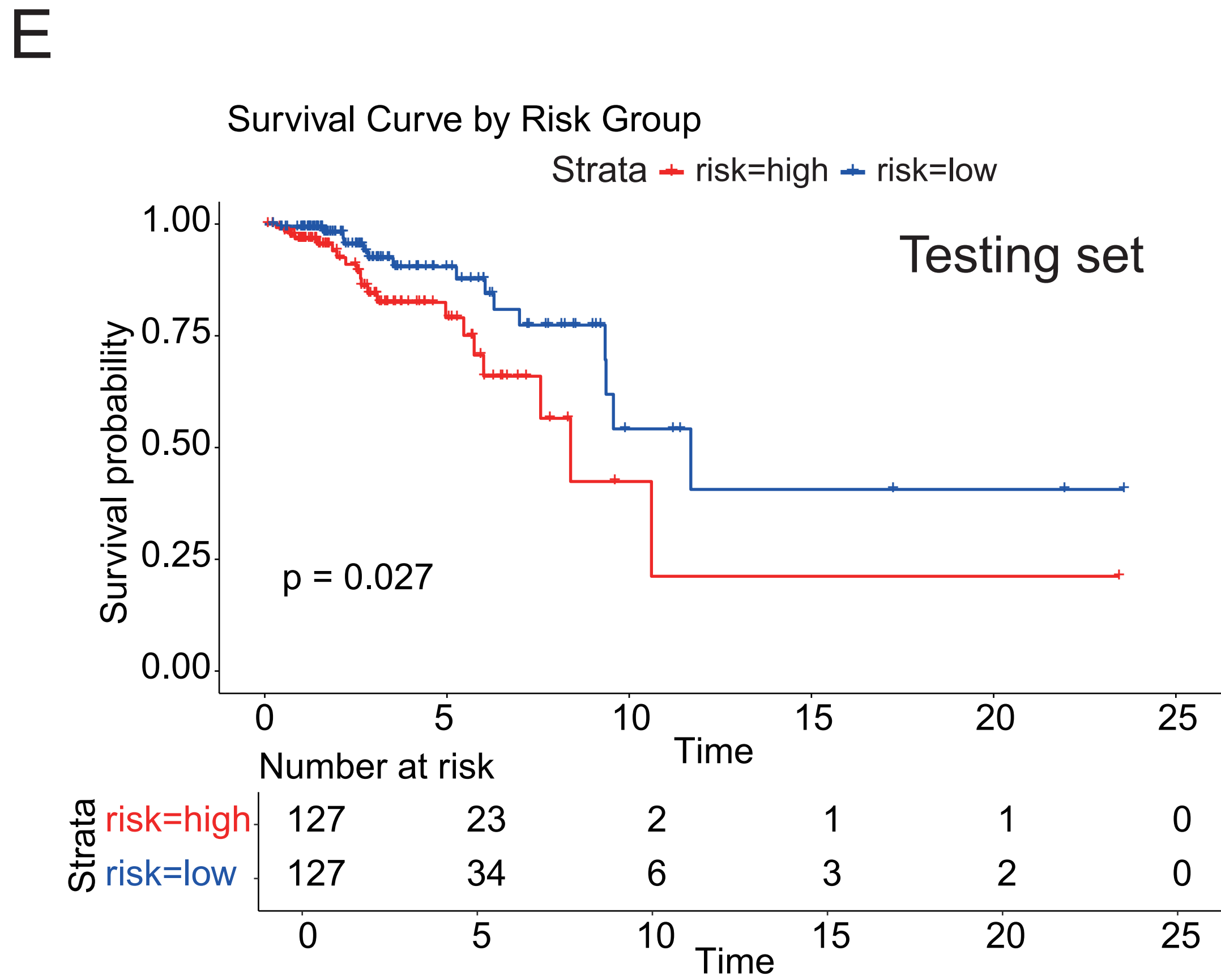
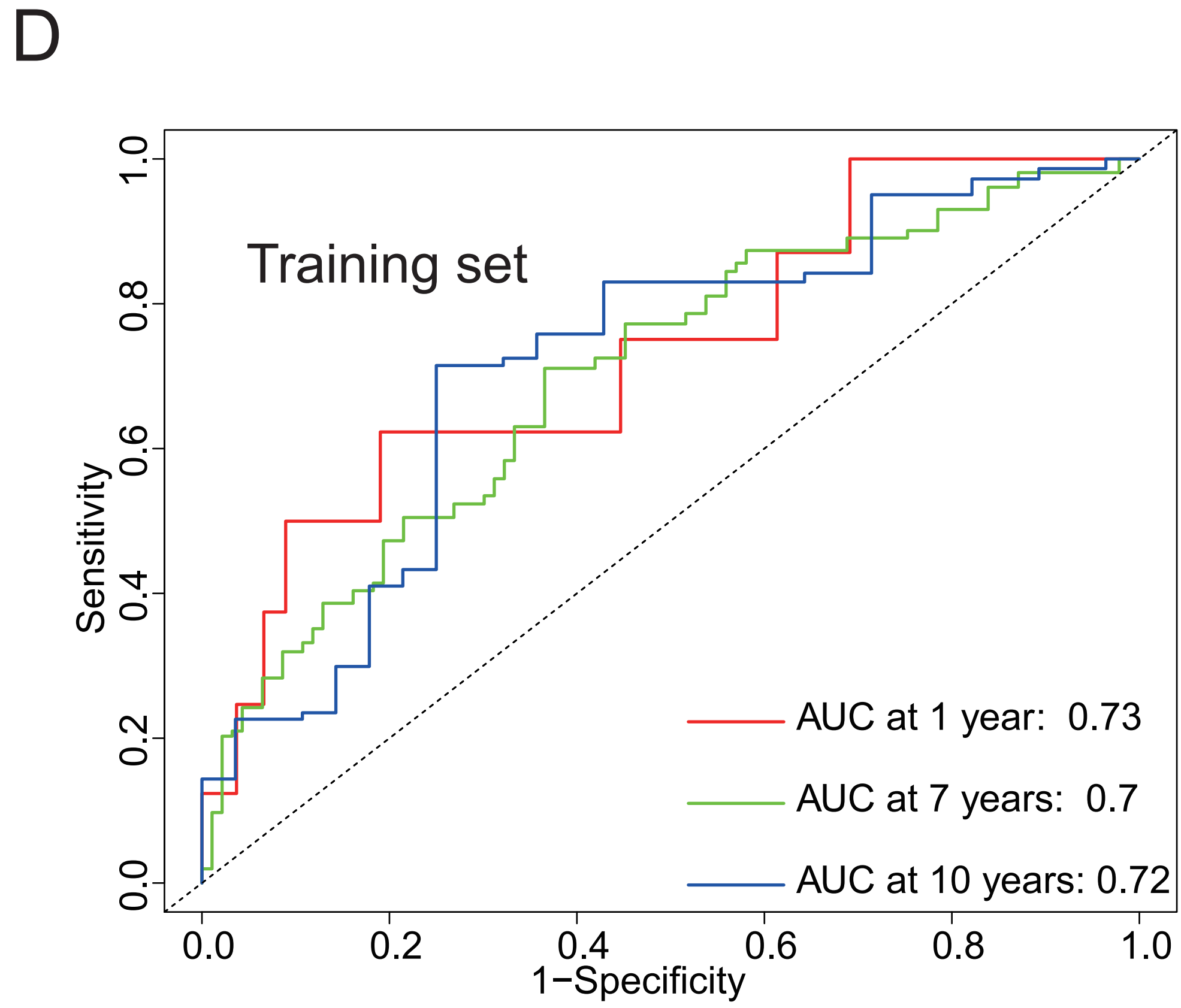
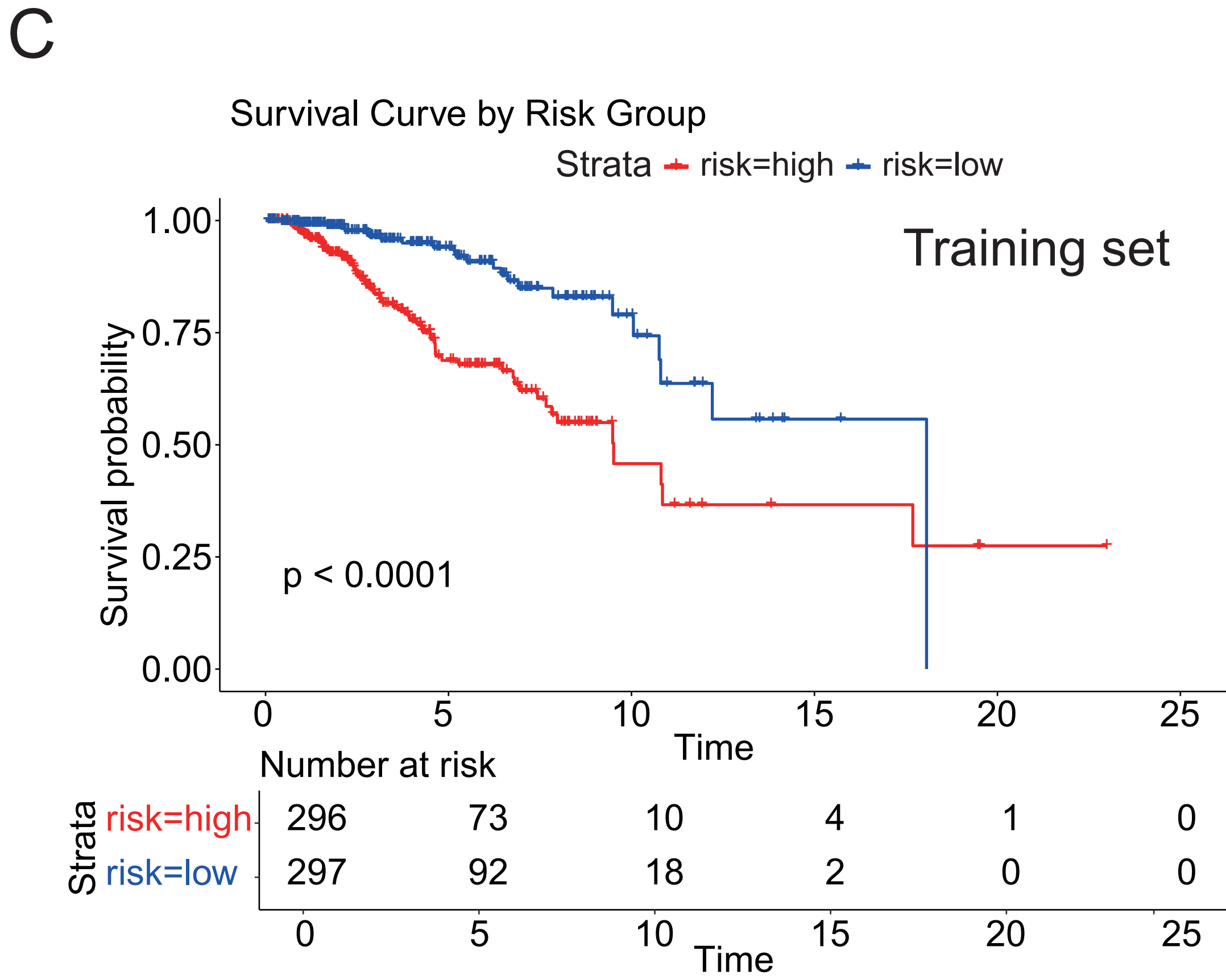
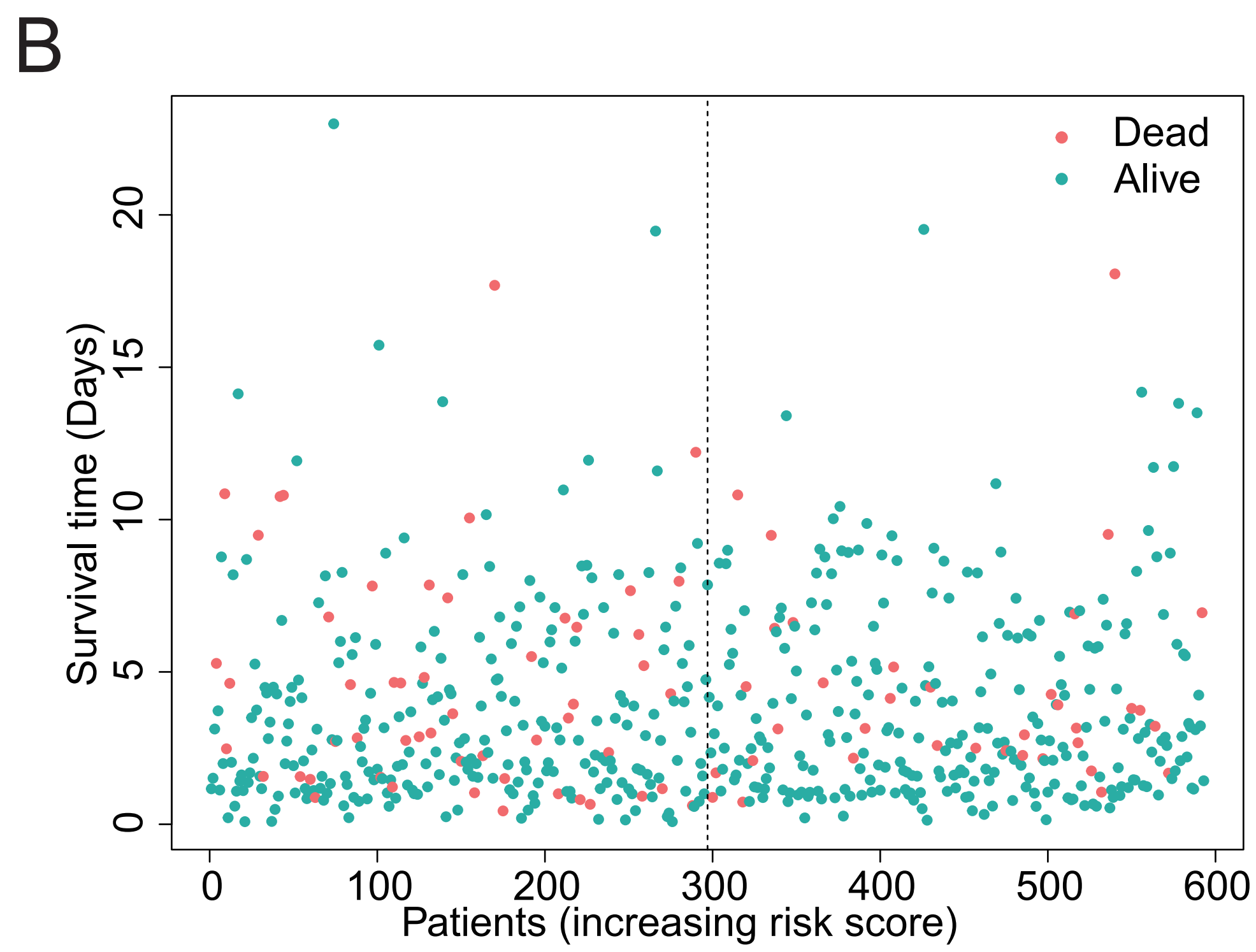
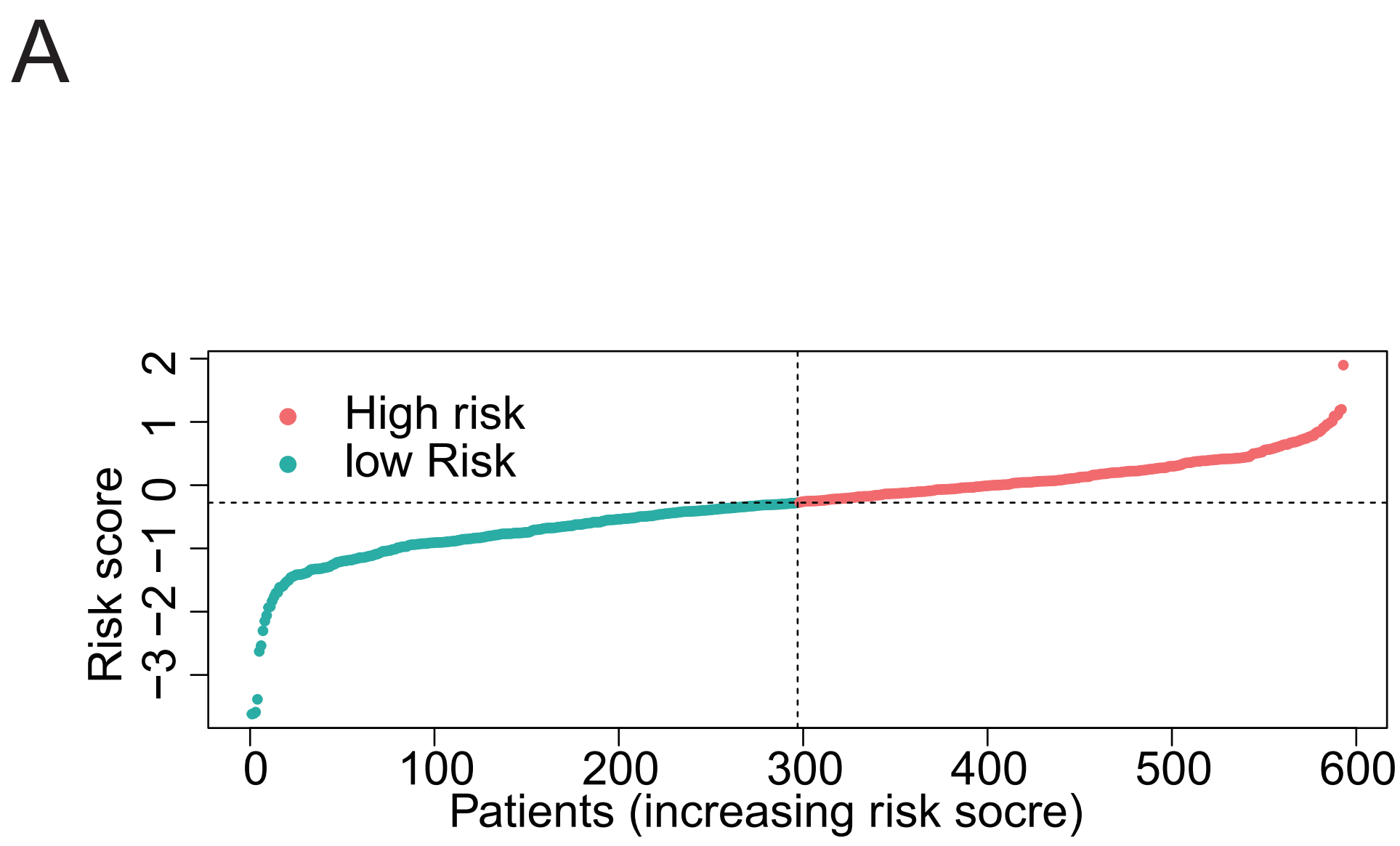
Hazard ratios

F

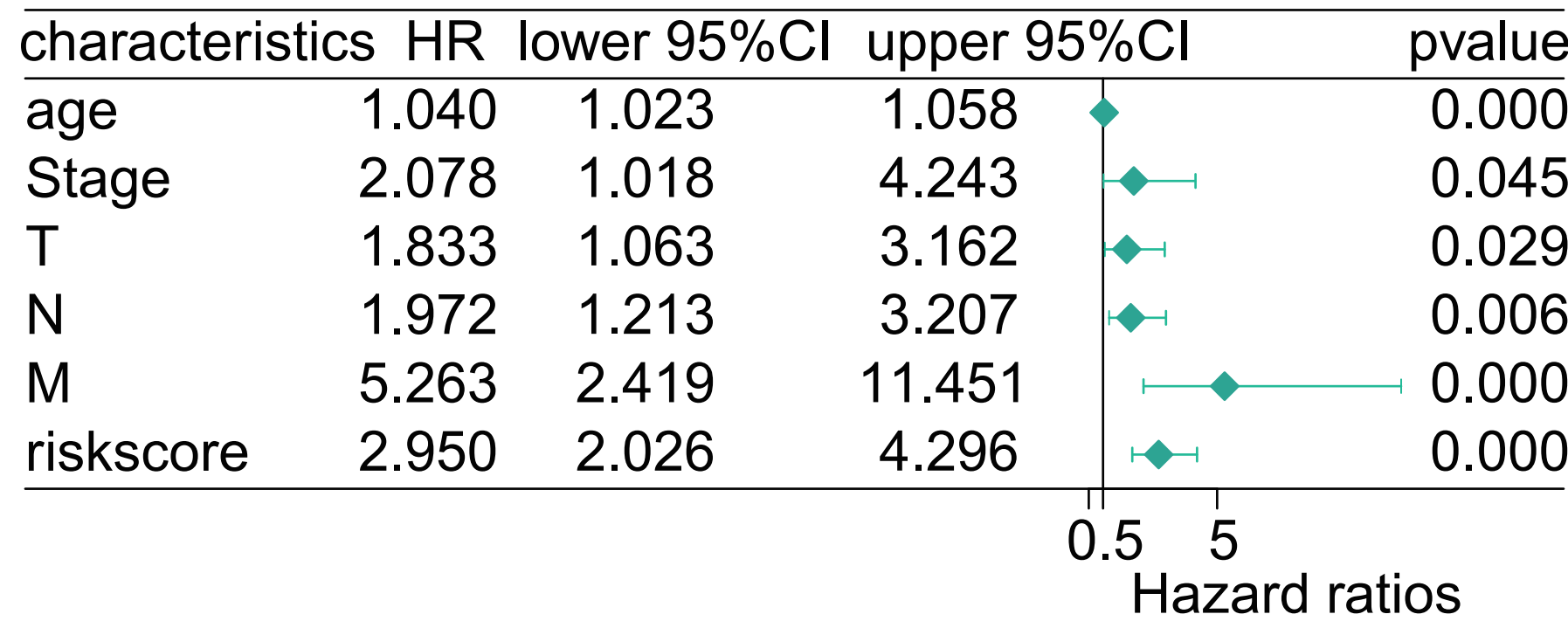


G

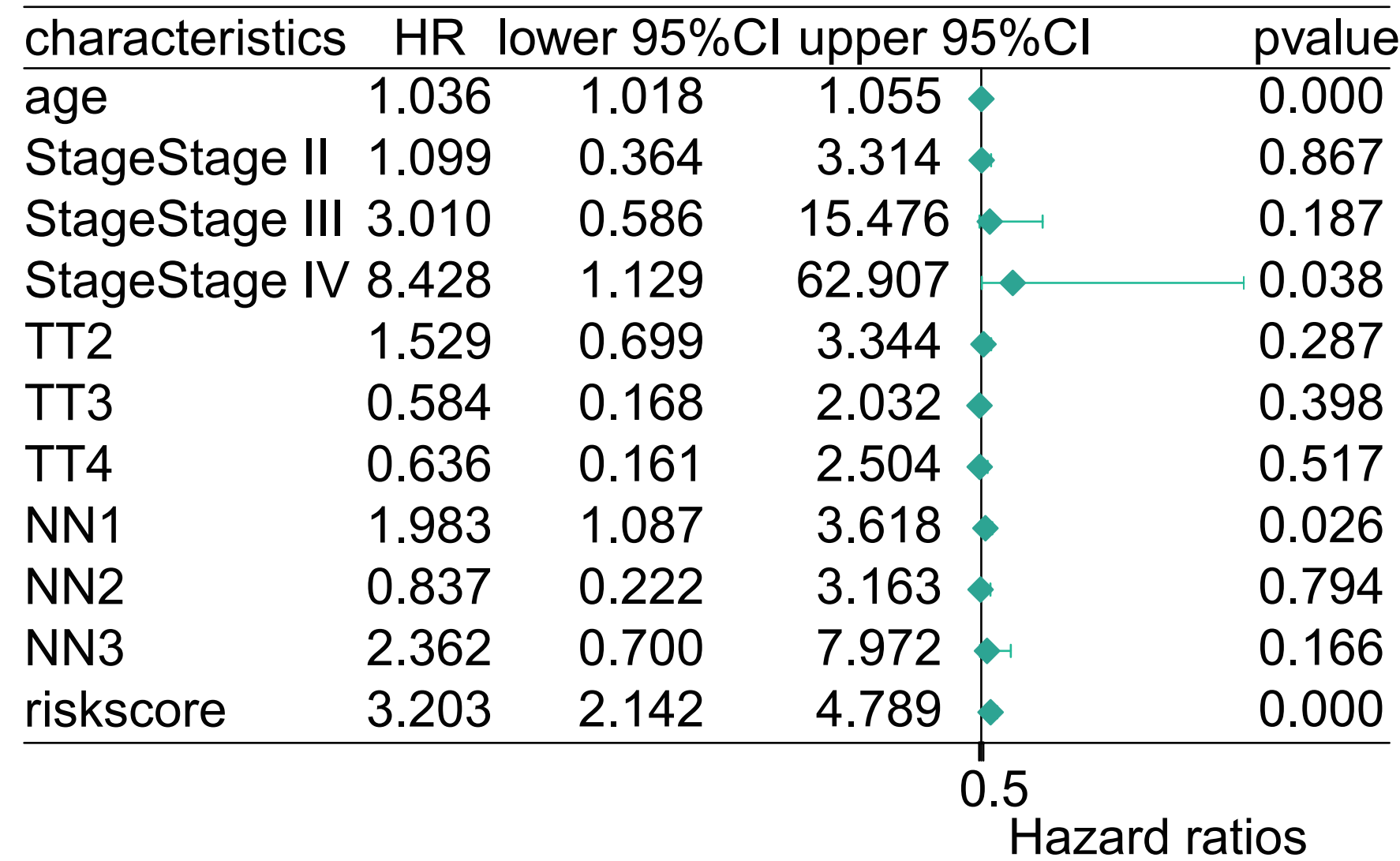




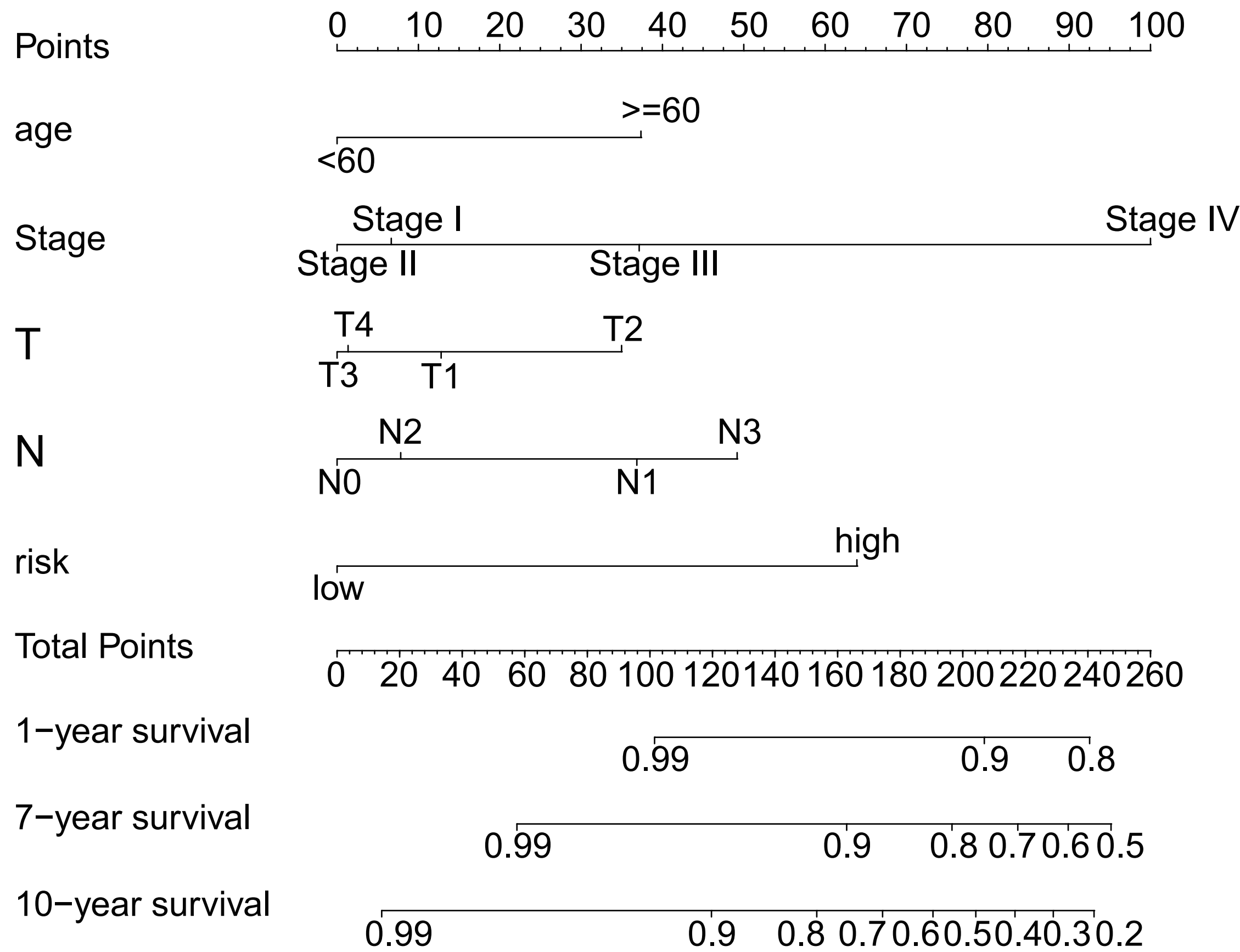
A



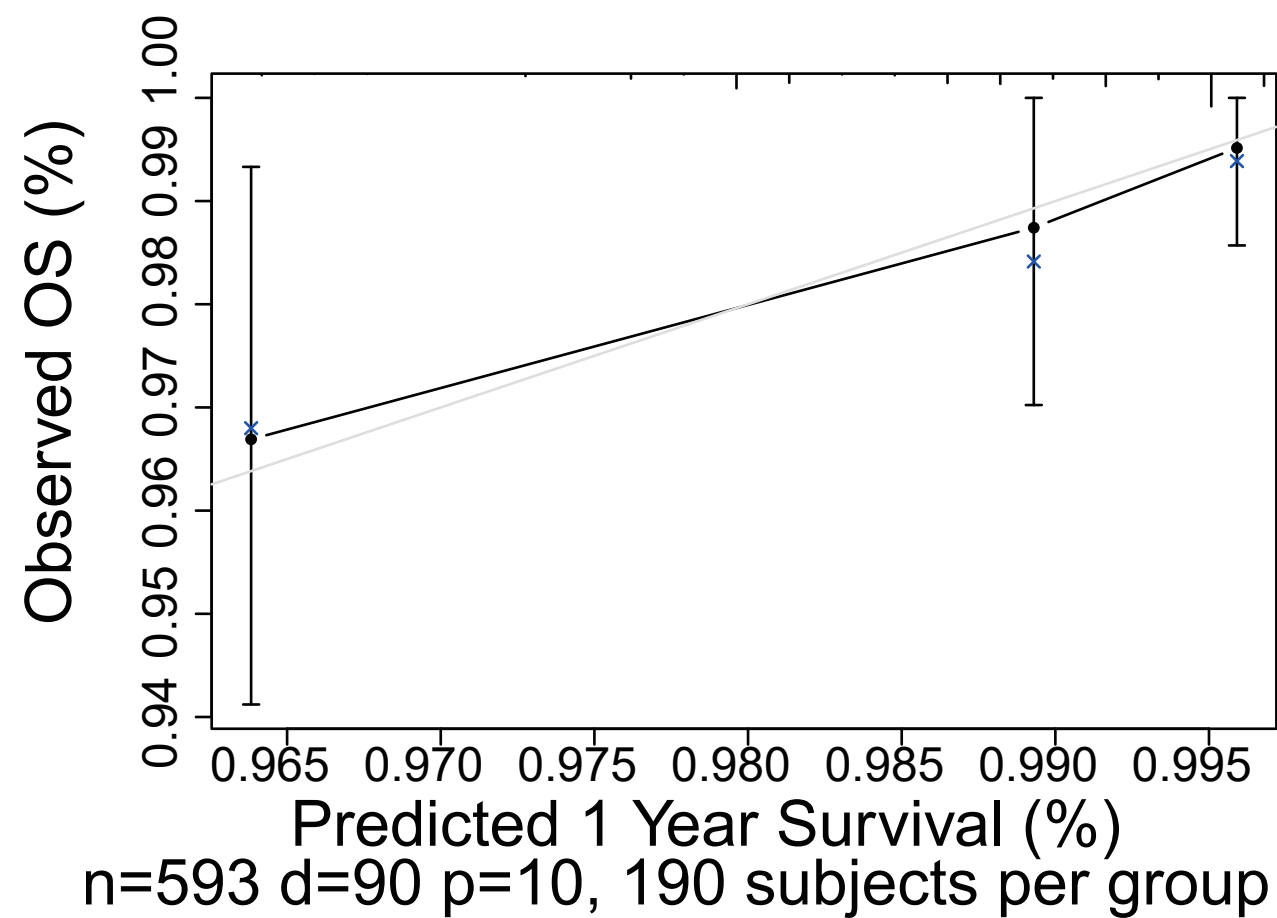
B



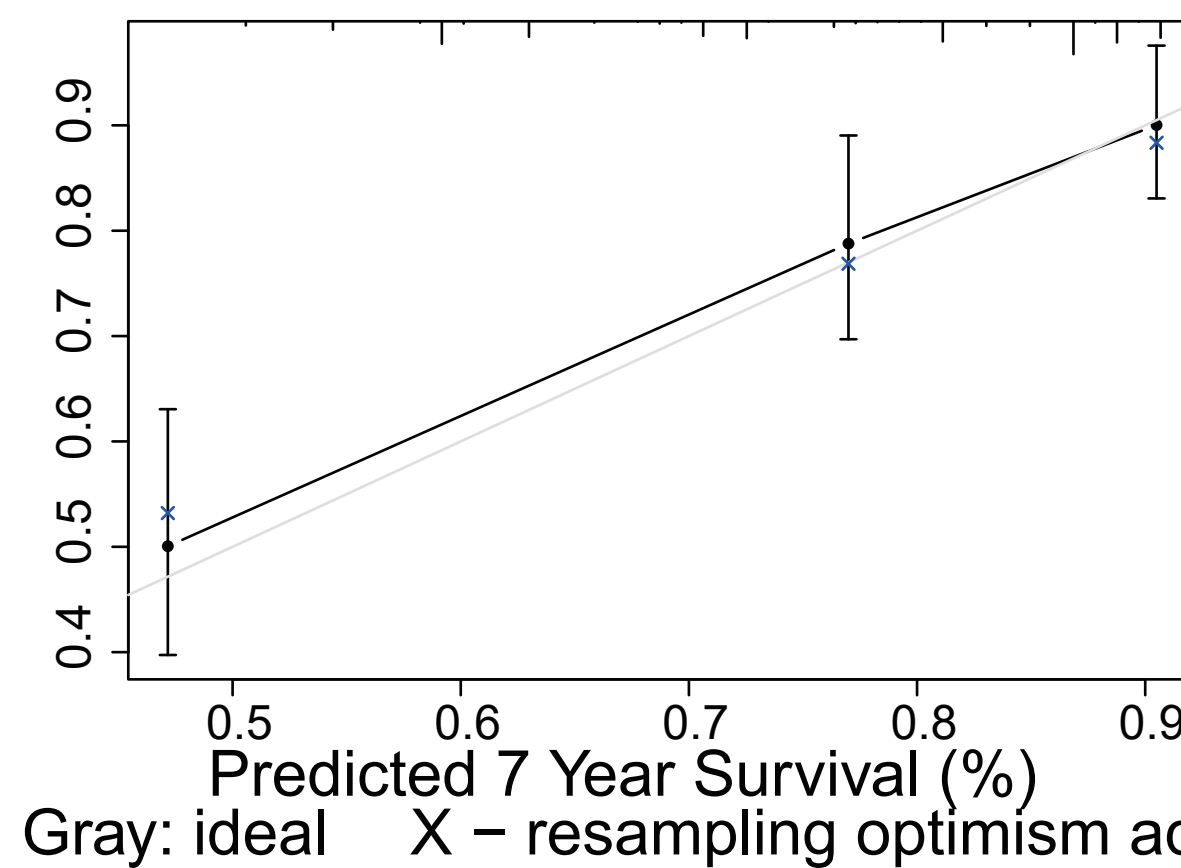
C



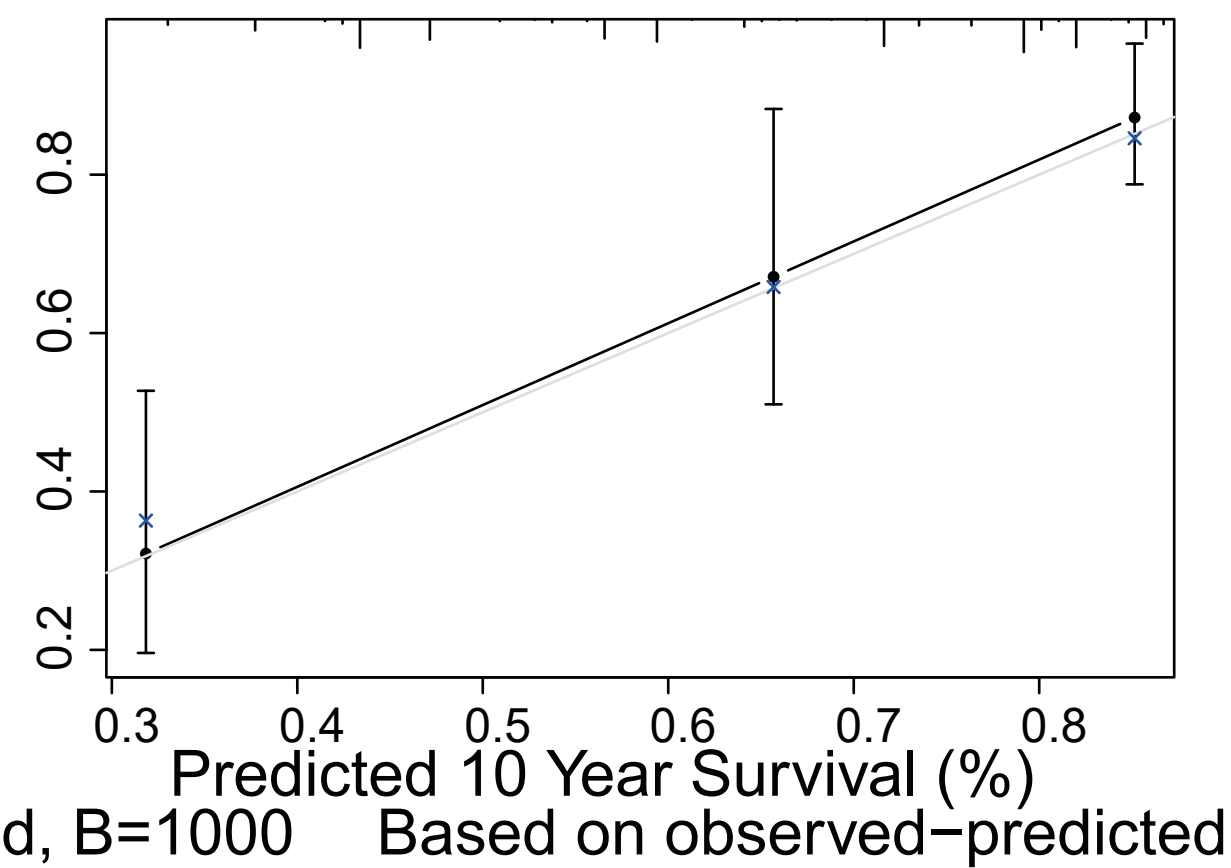
D



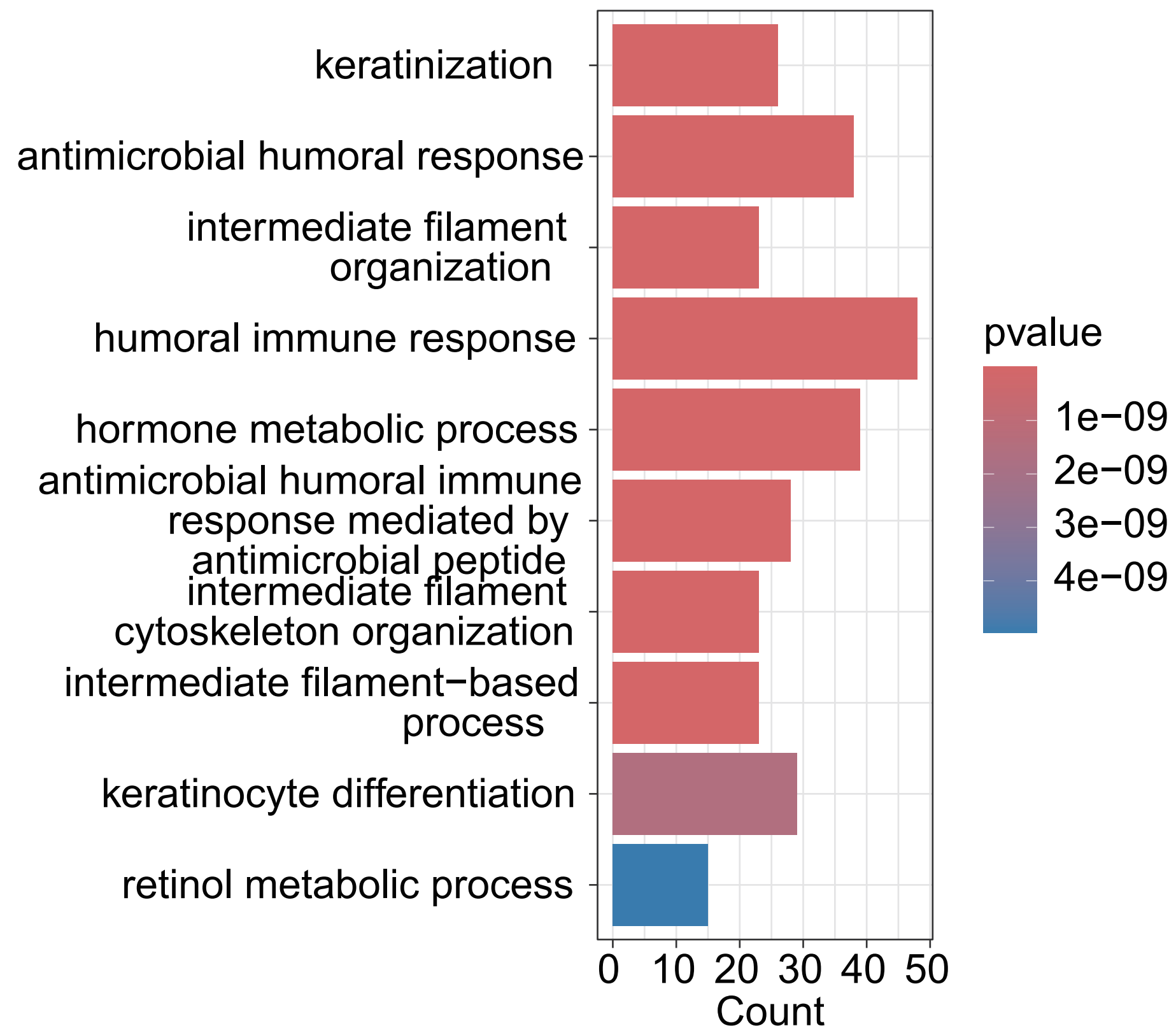
E



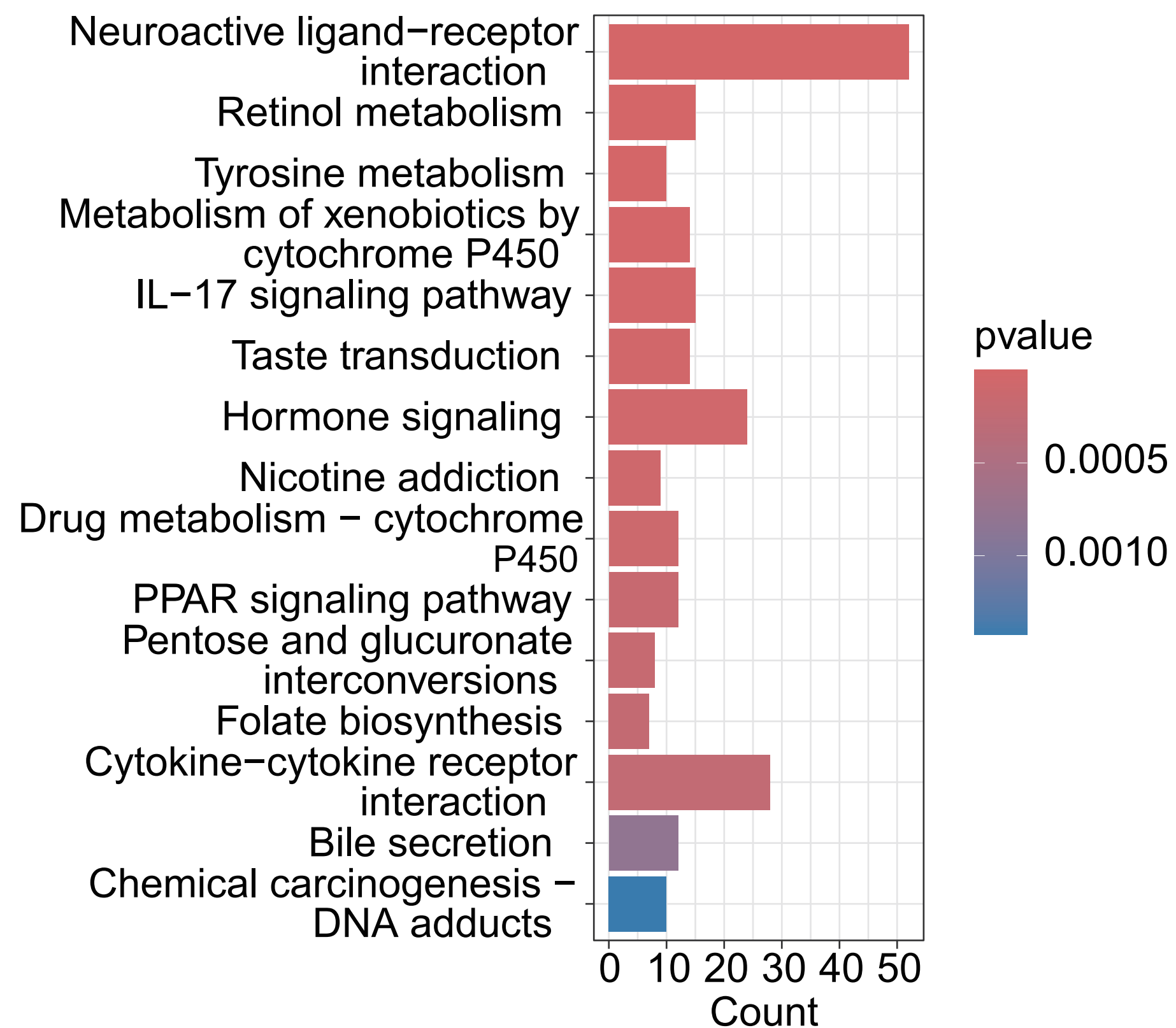
F



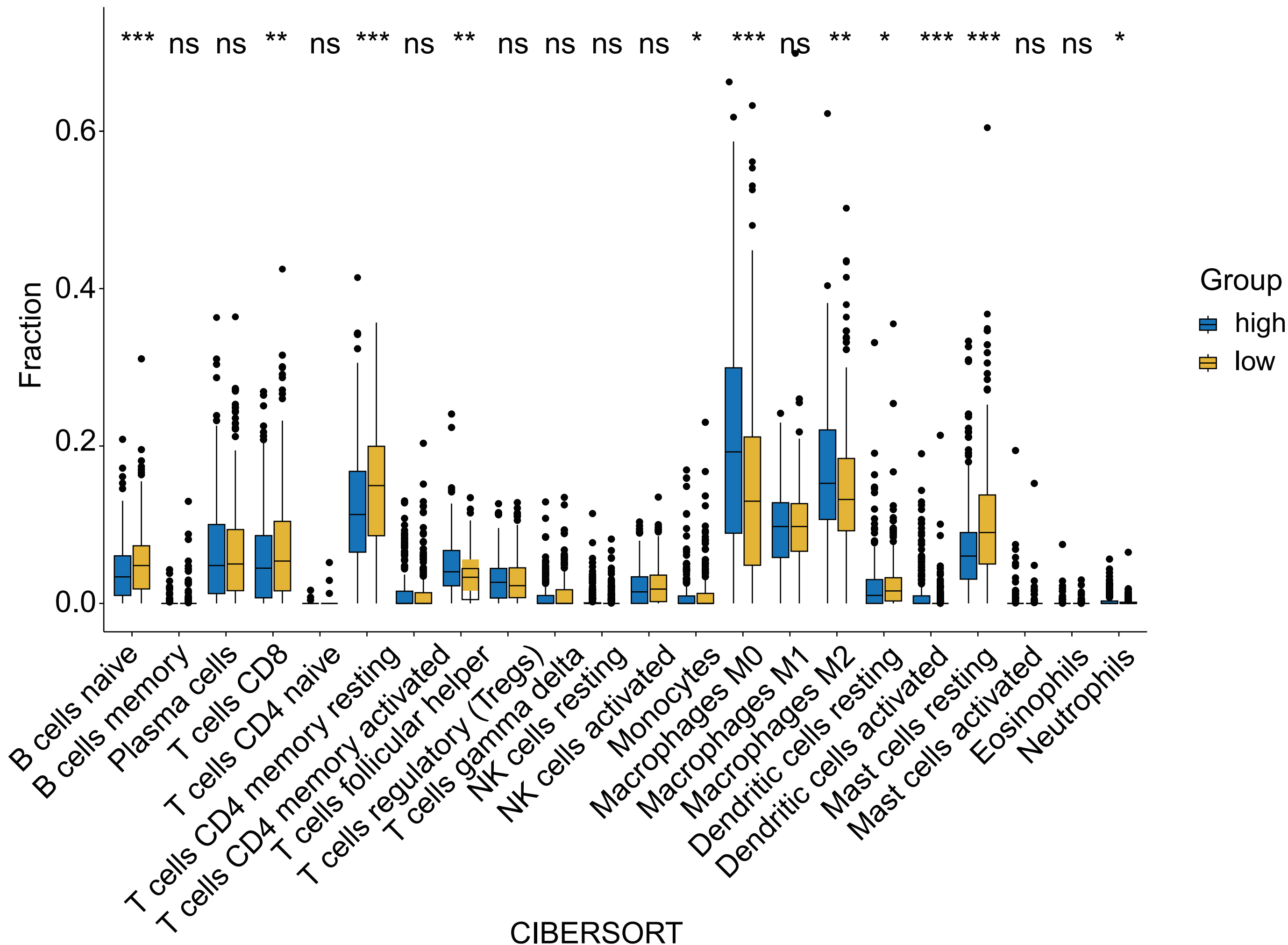
A



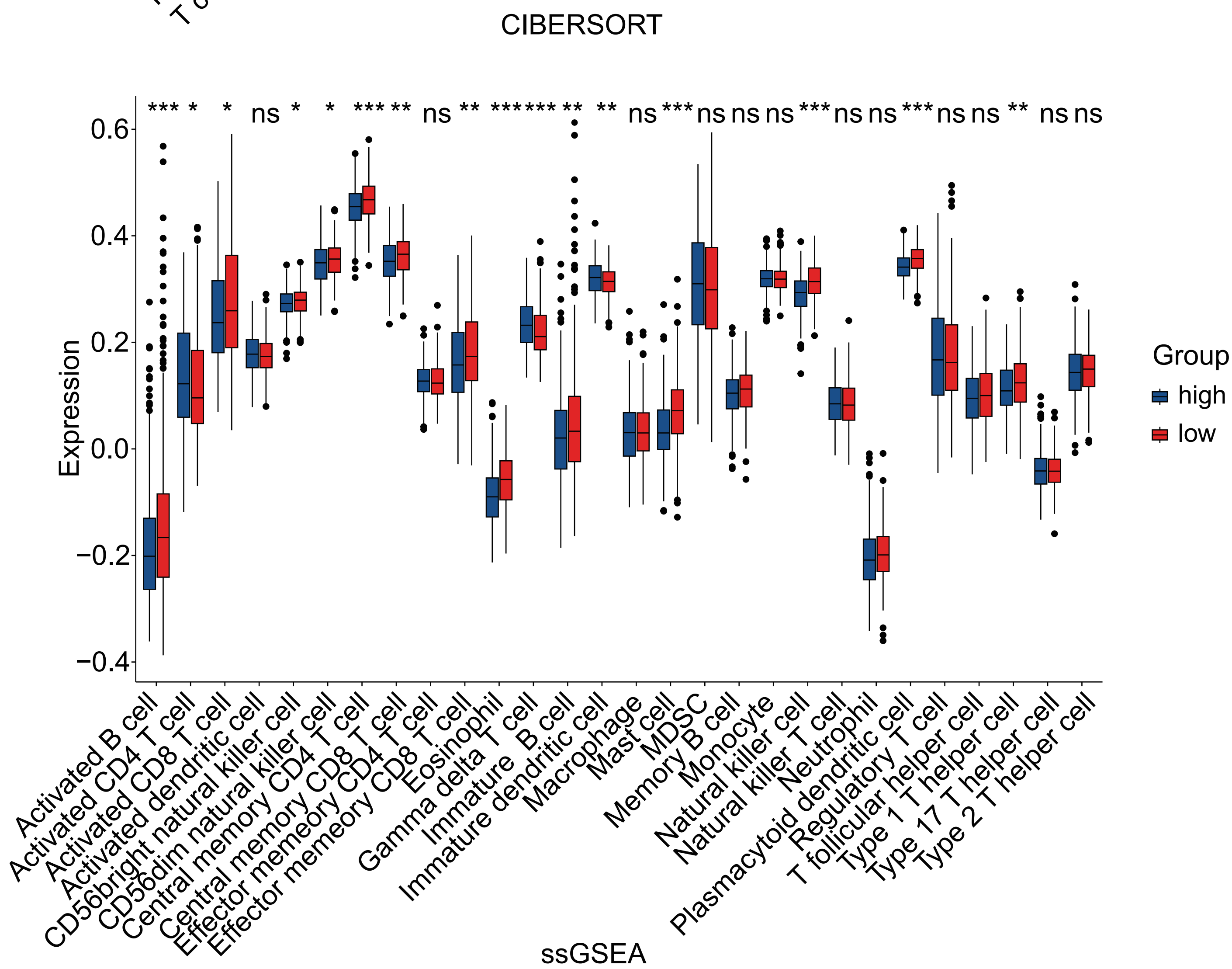
B



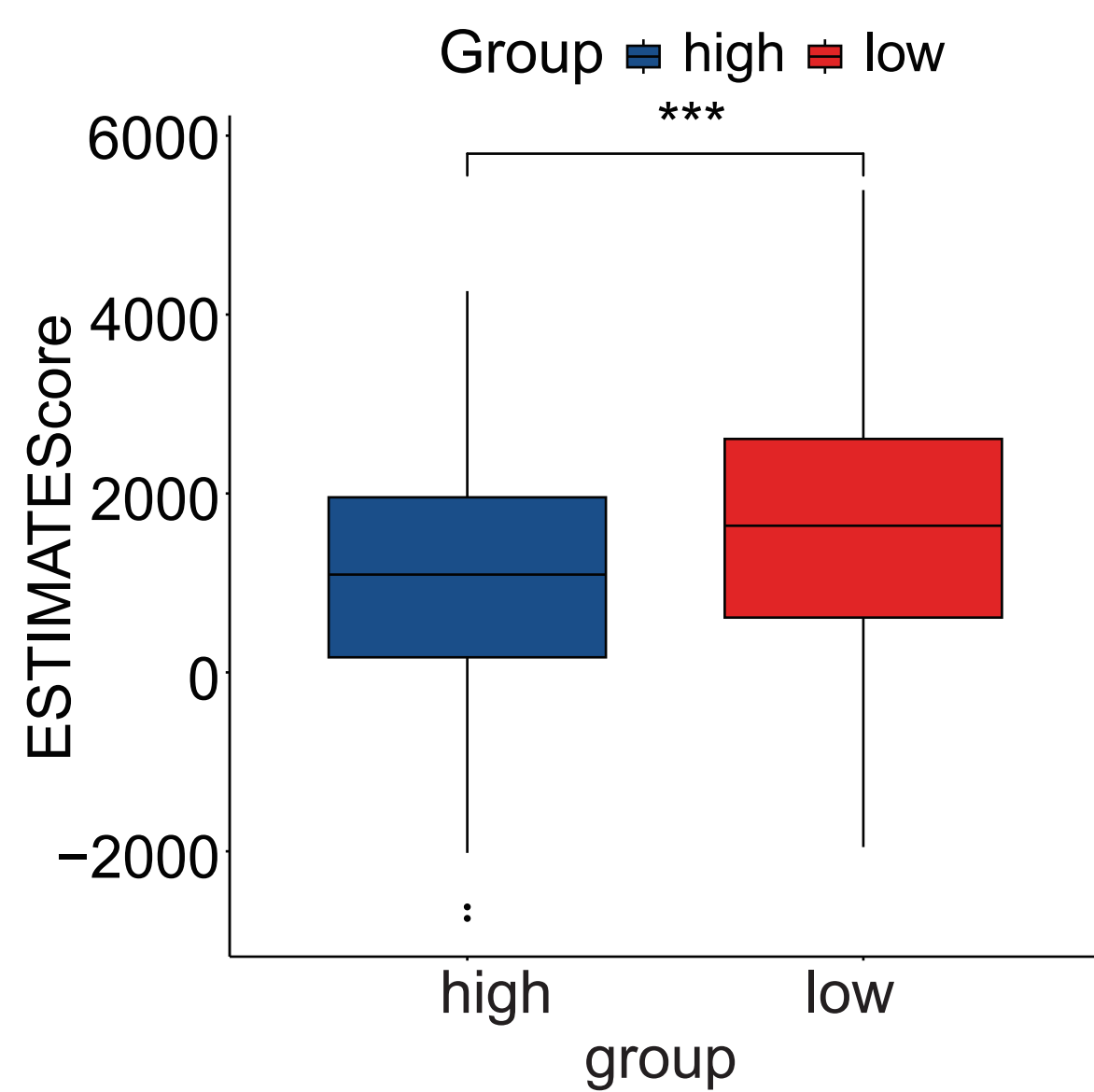
A



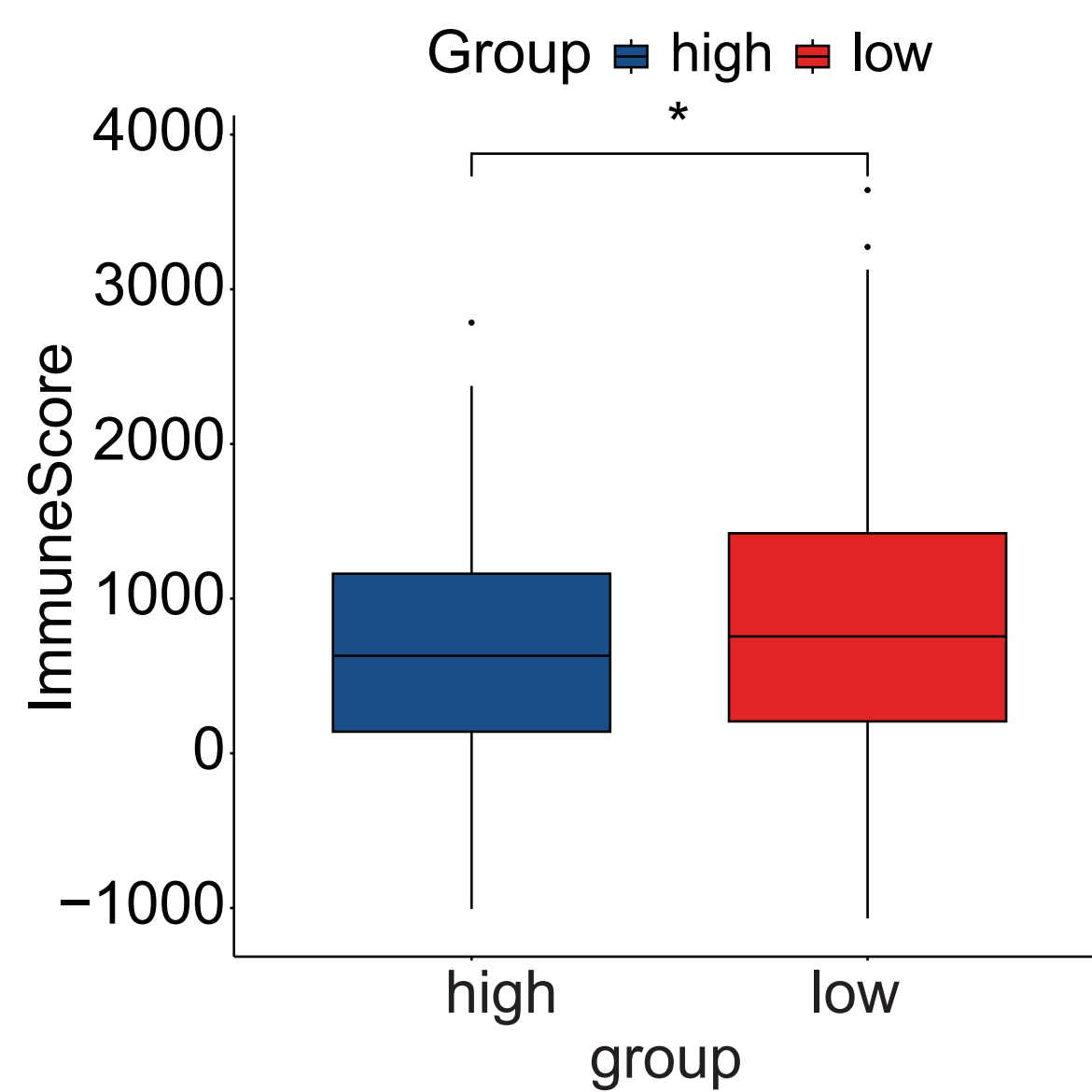
B



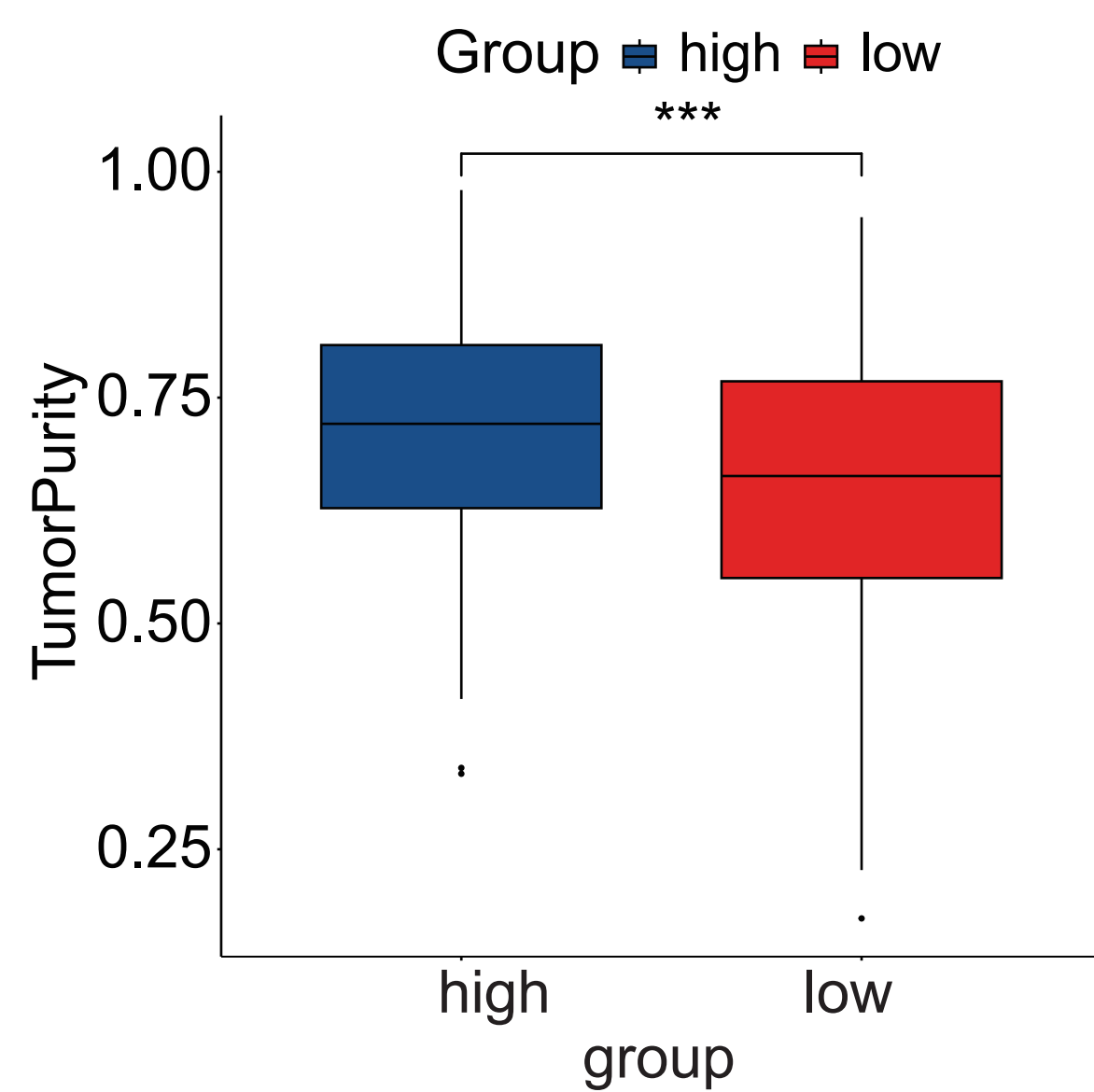
C



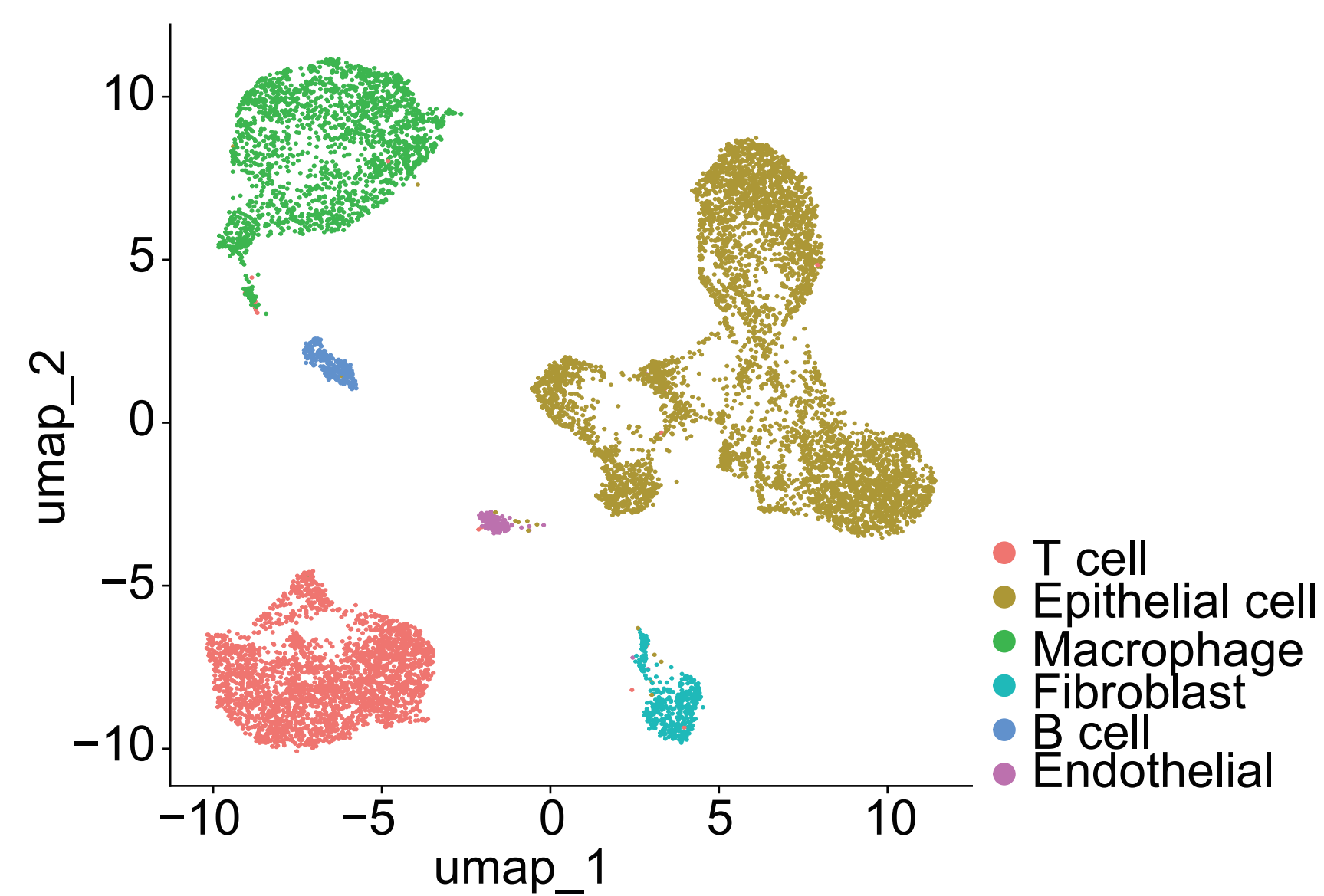
D



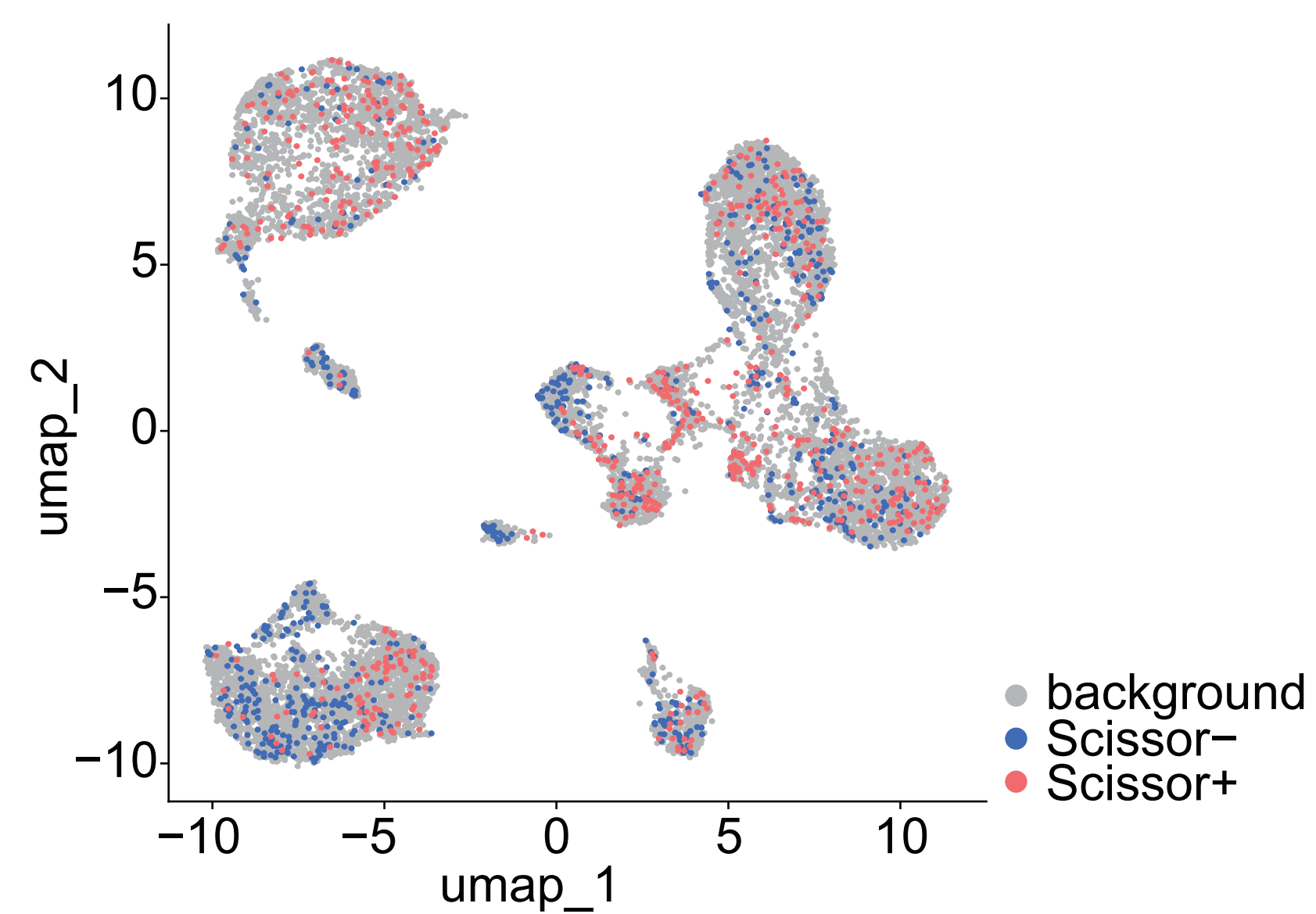
E



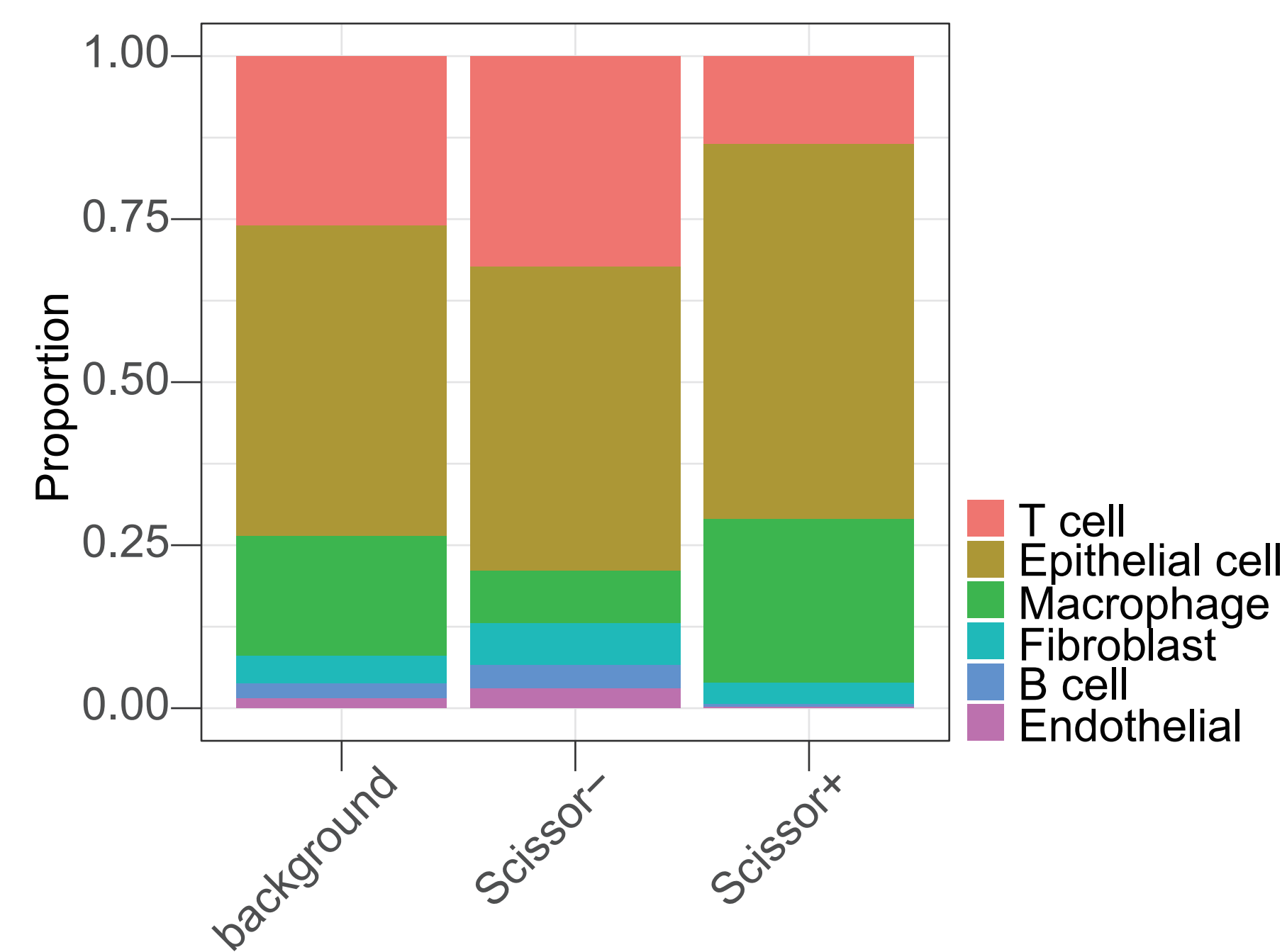
A



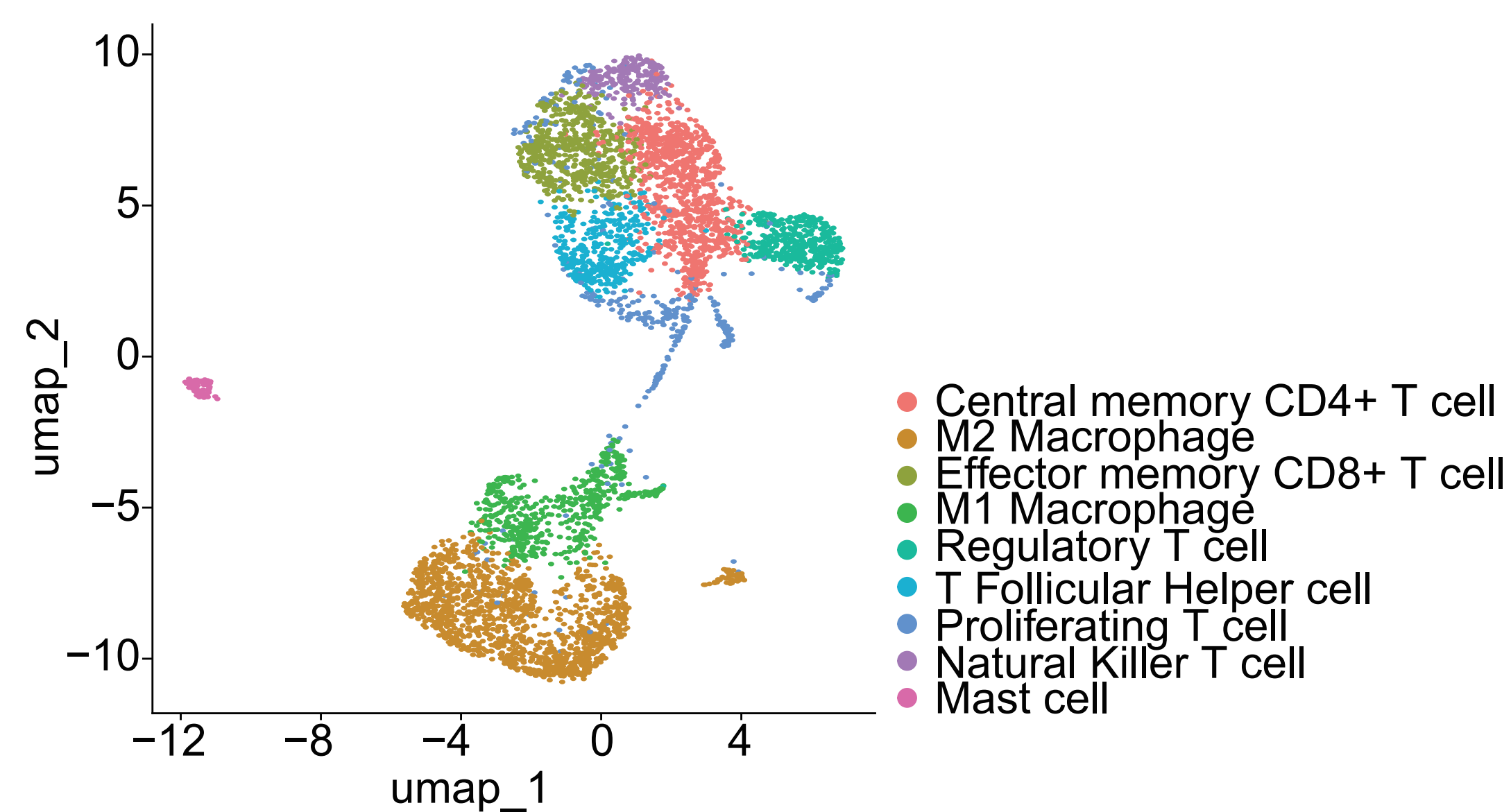
B



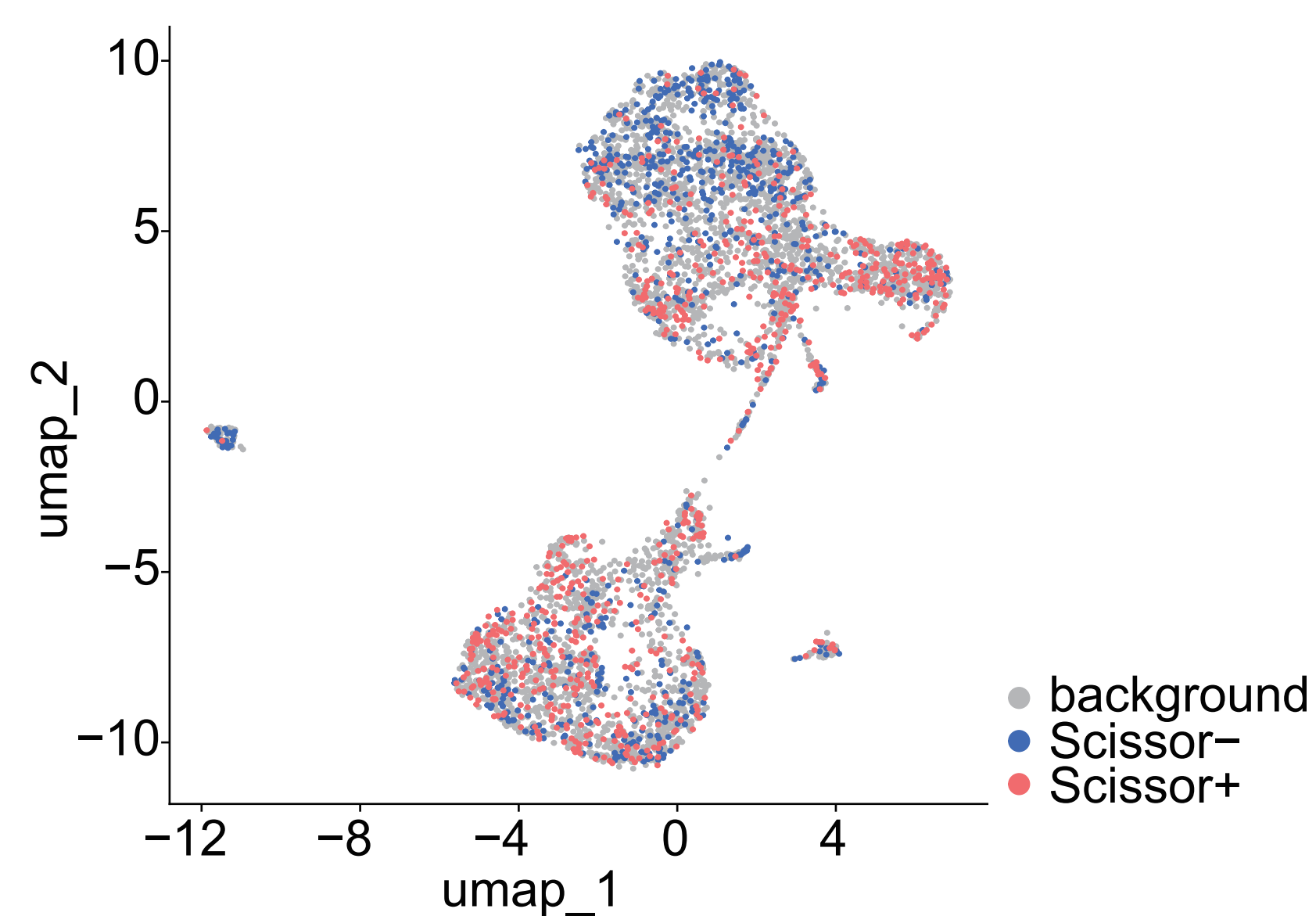
C



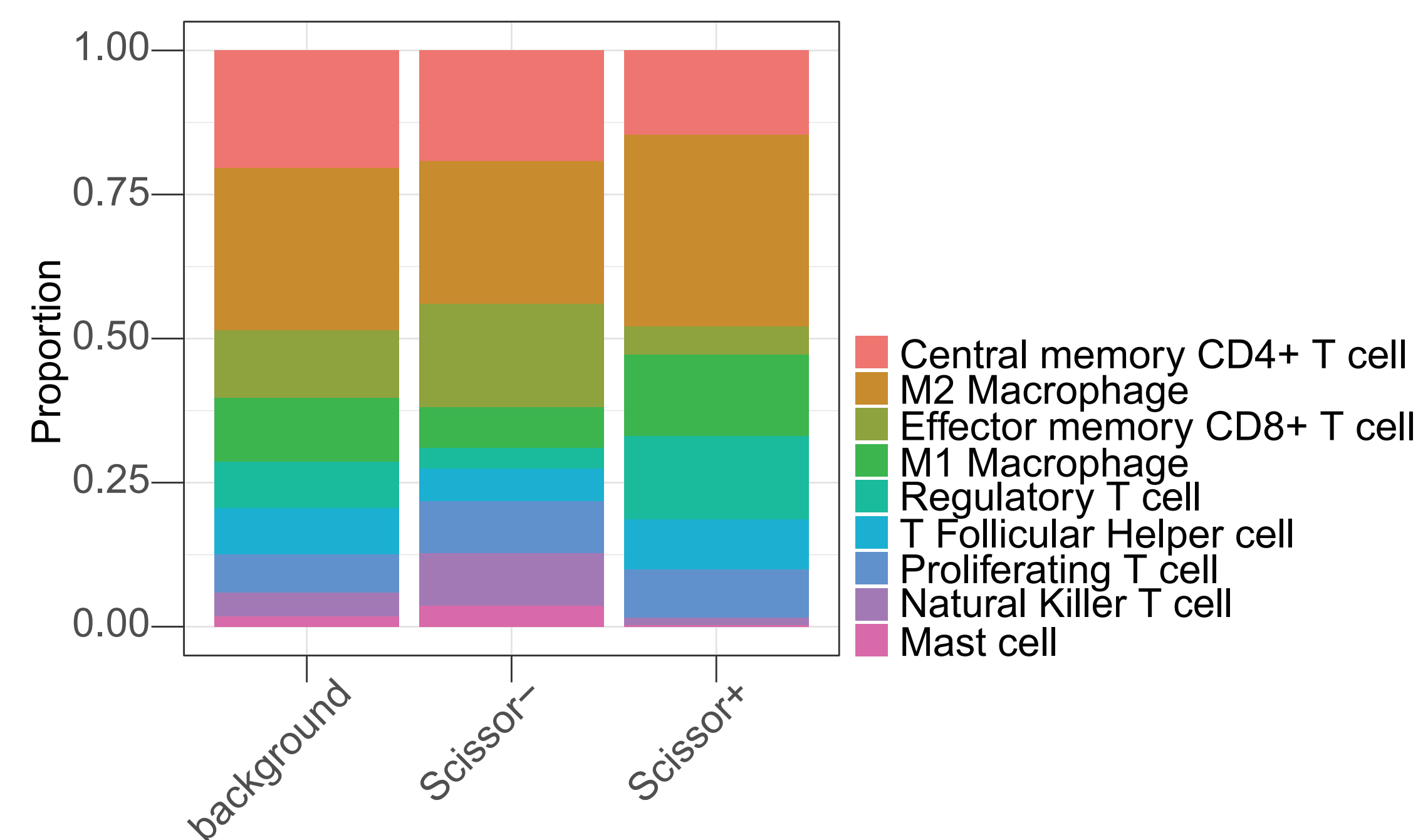
D



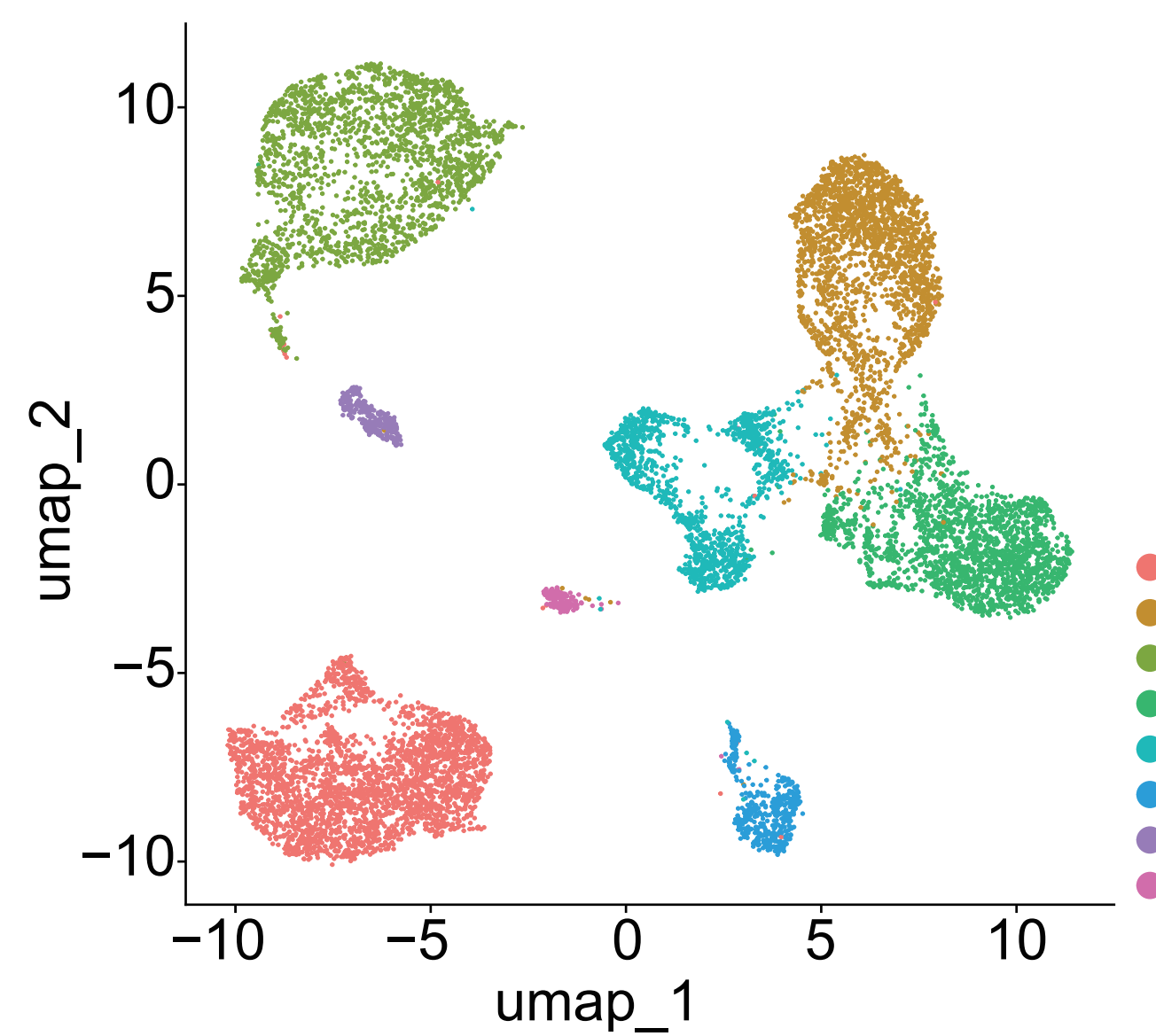
E



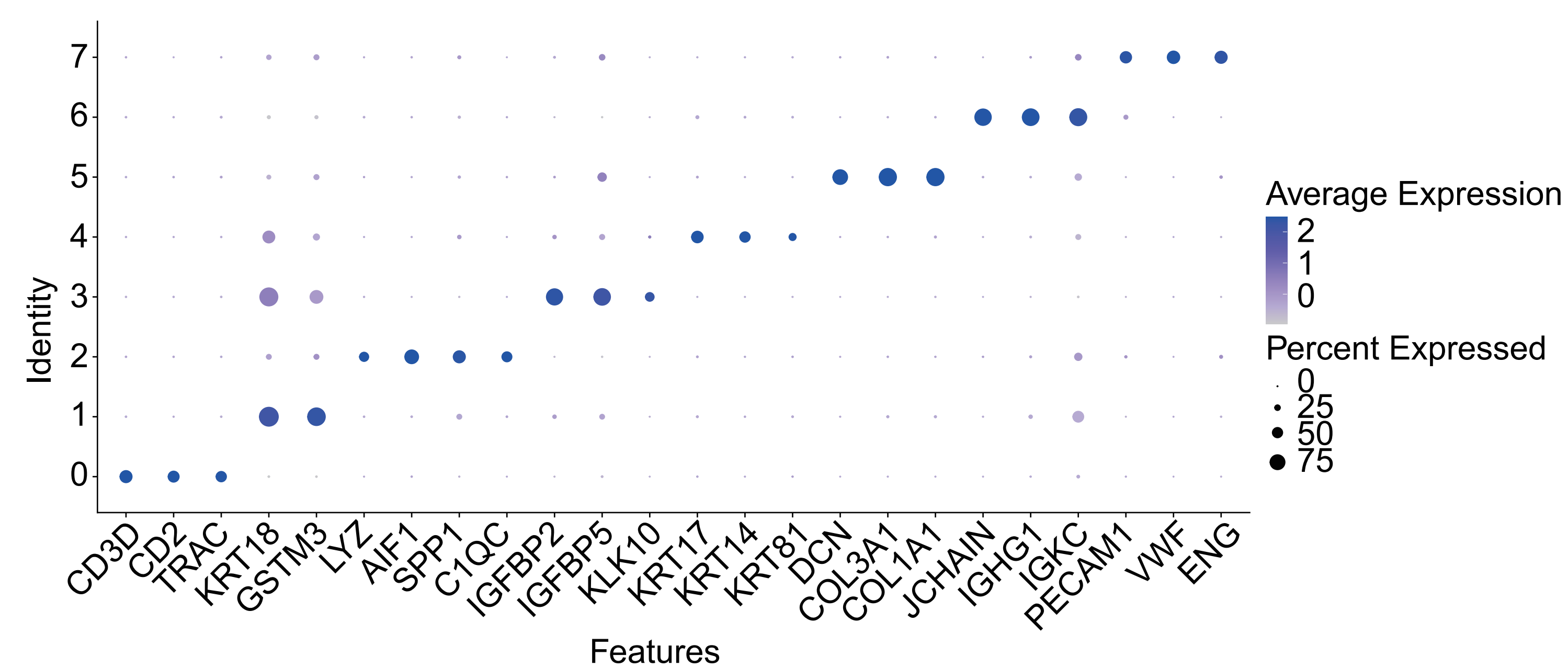
F



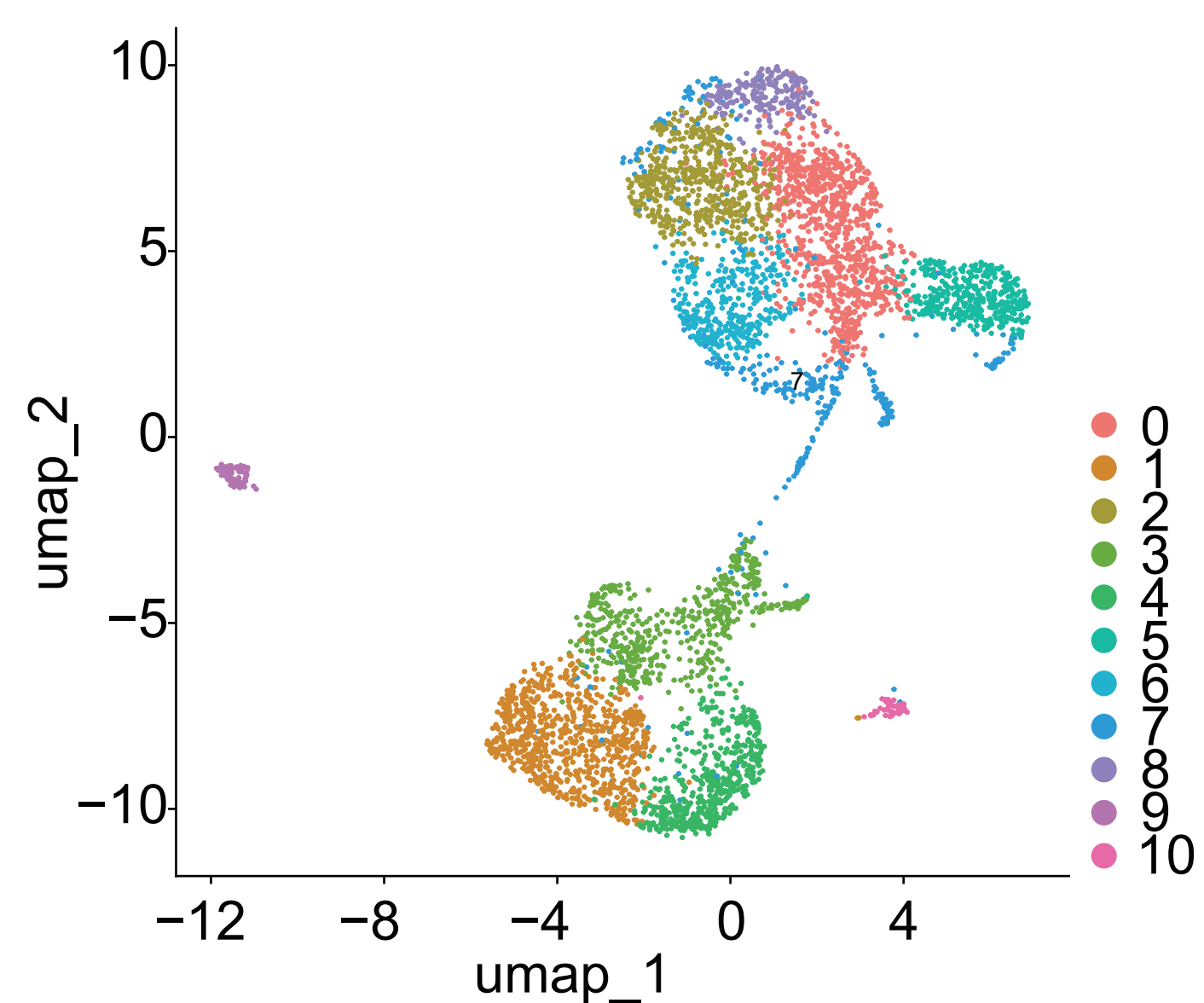
G



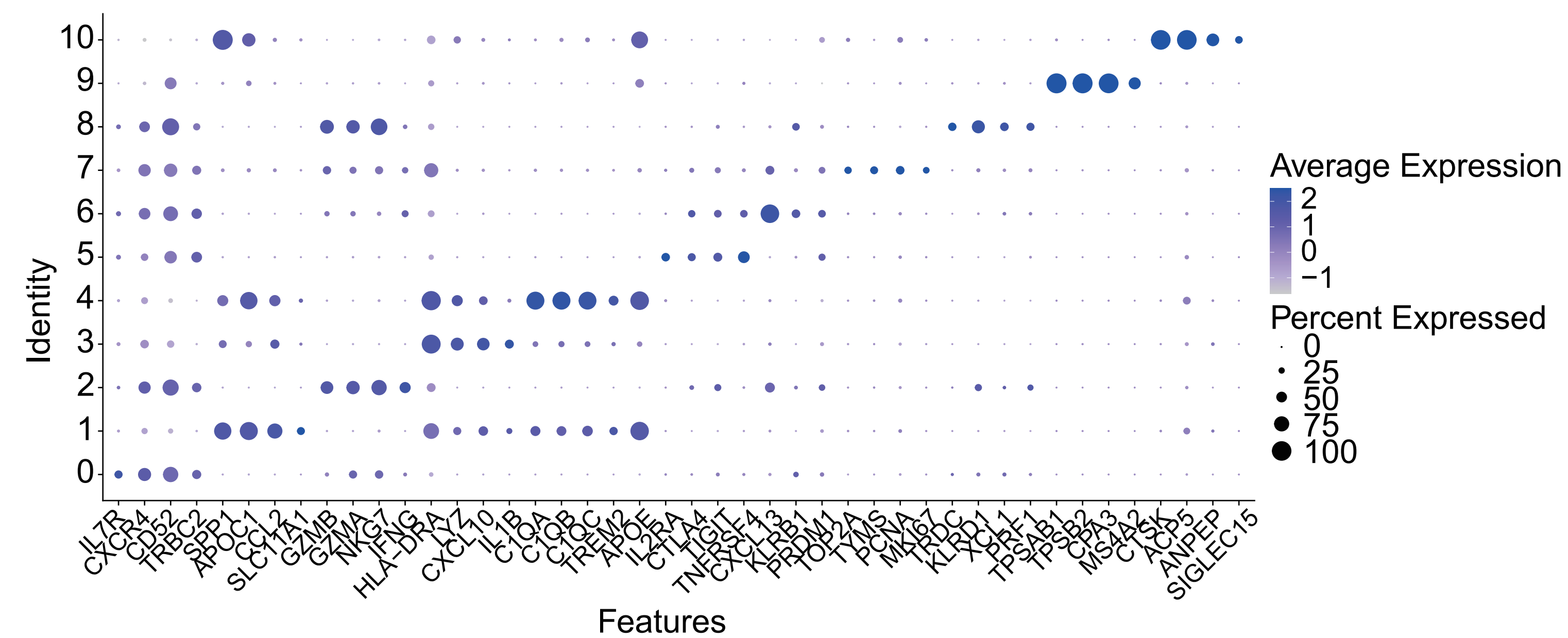
H



I



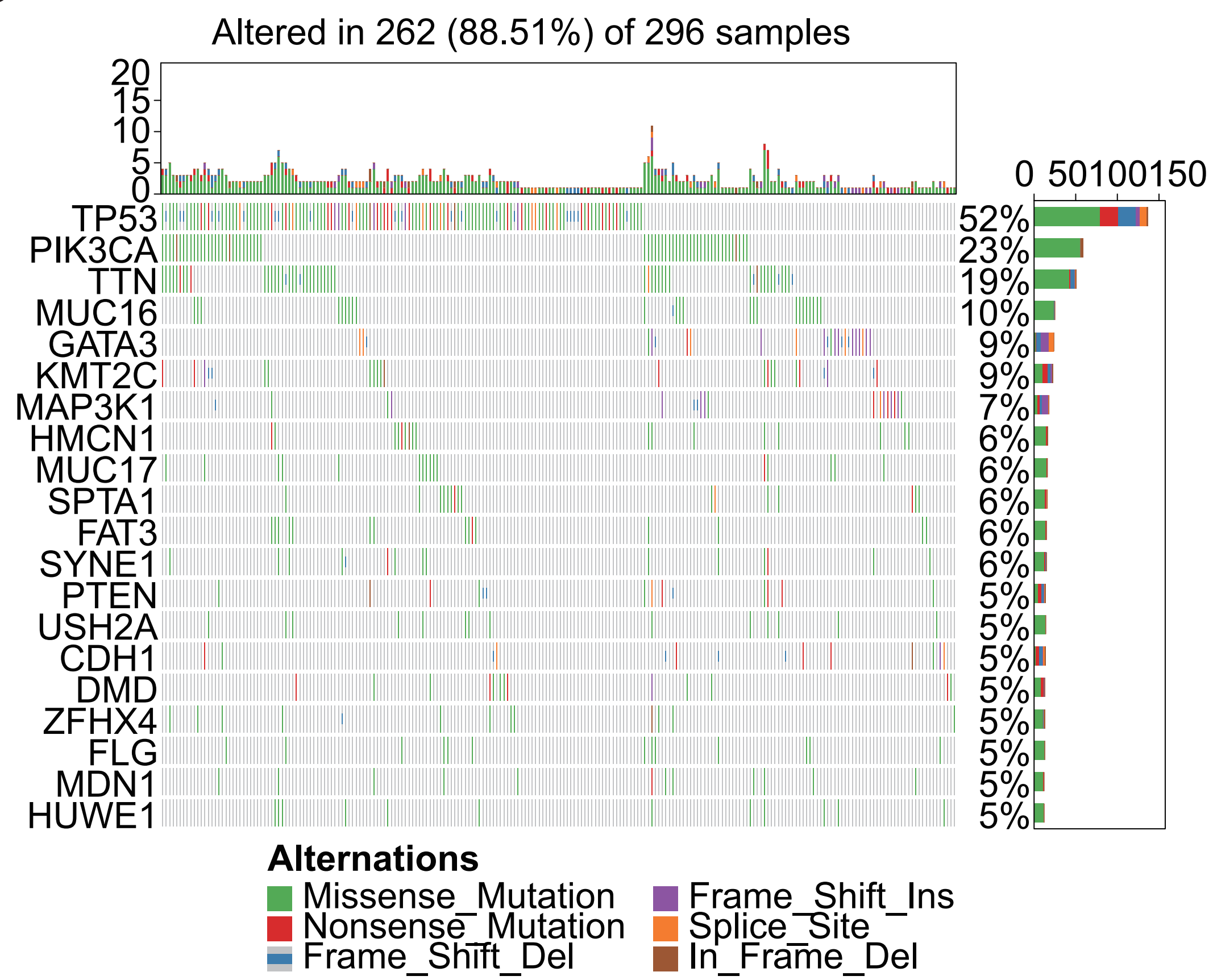
J



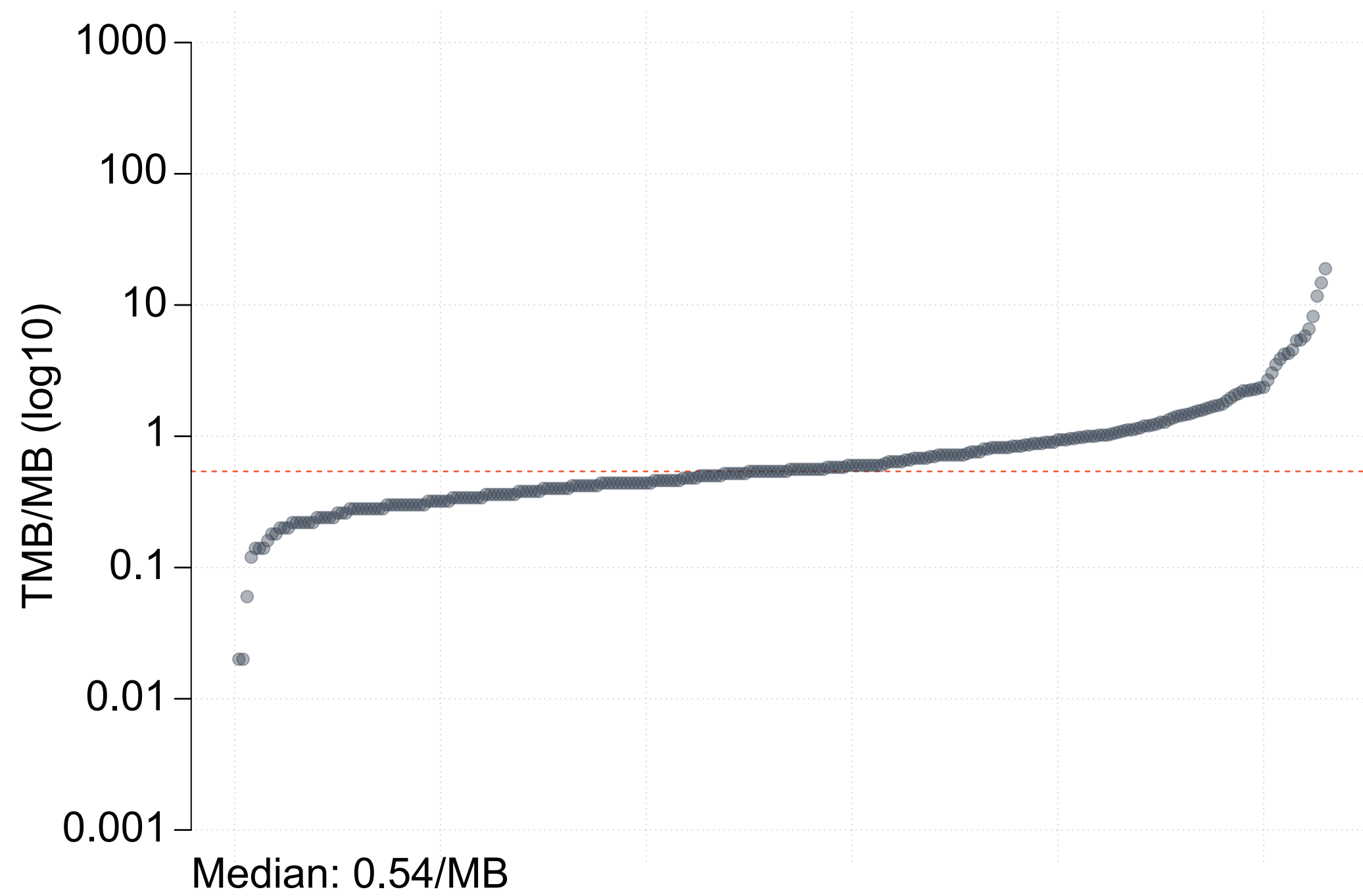
A



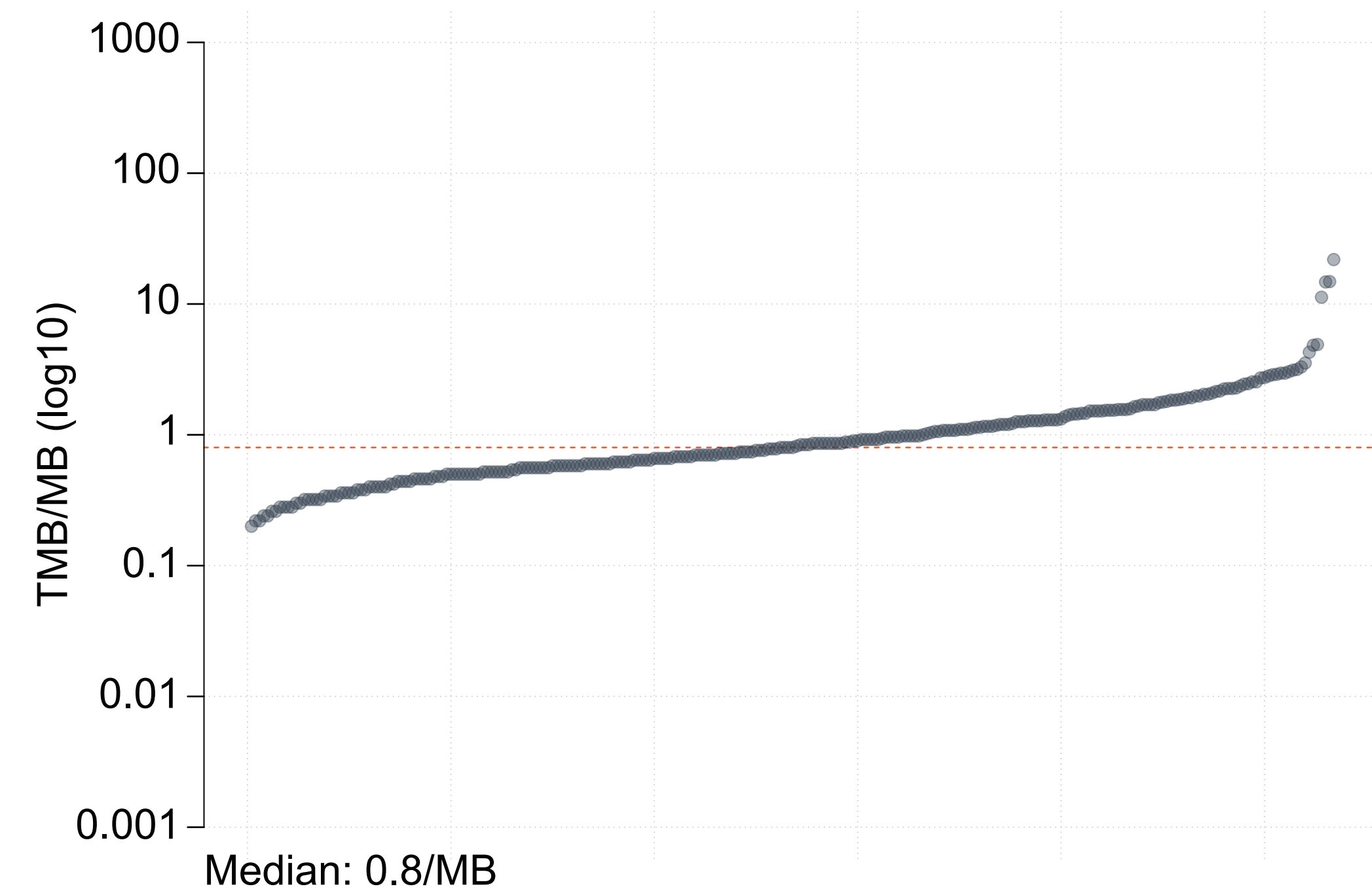
B



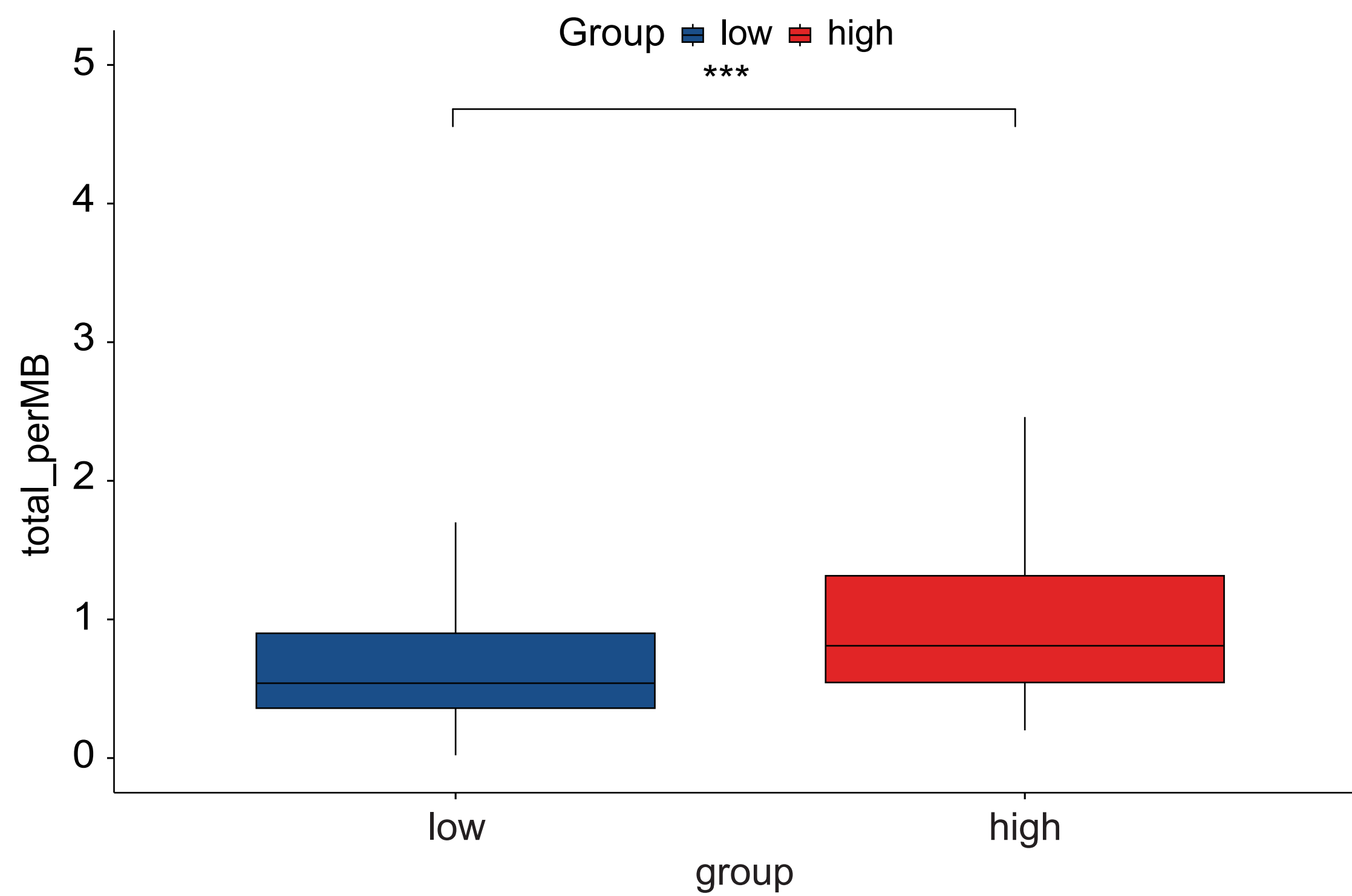
C

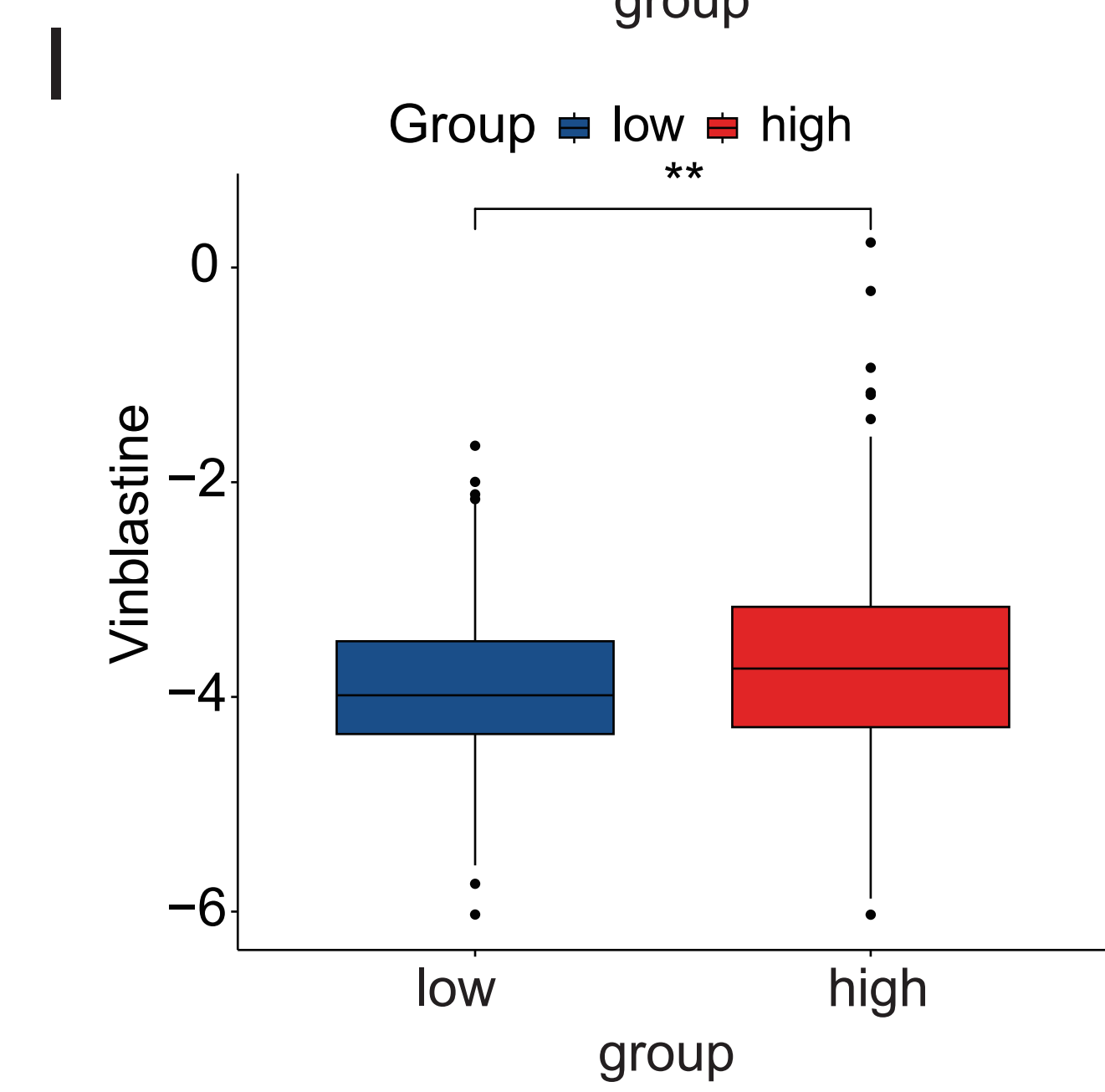
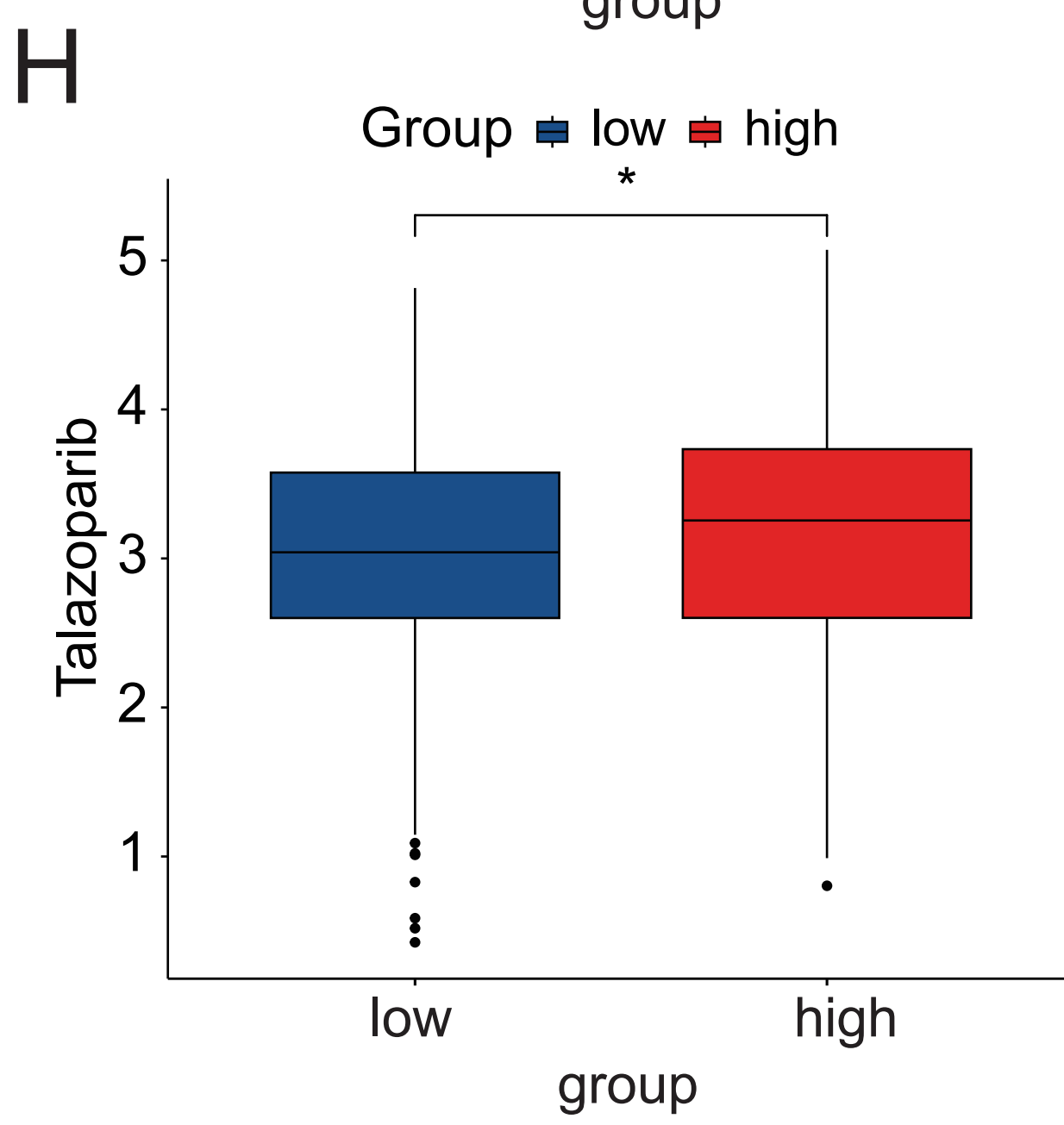
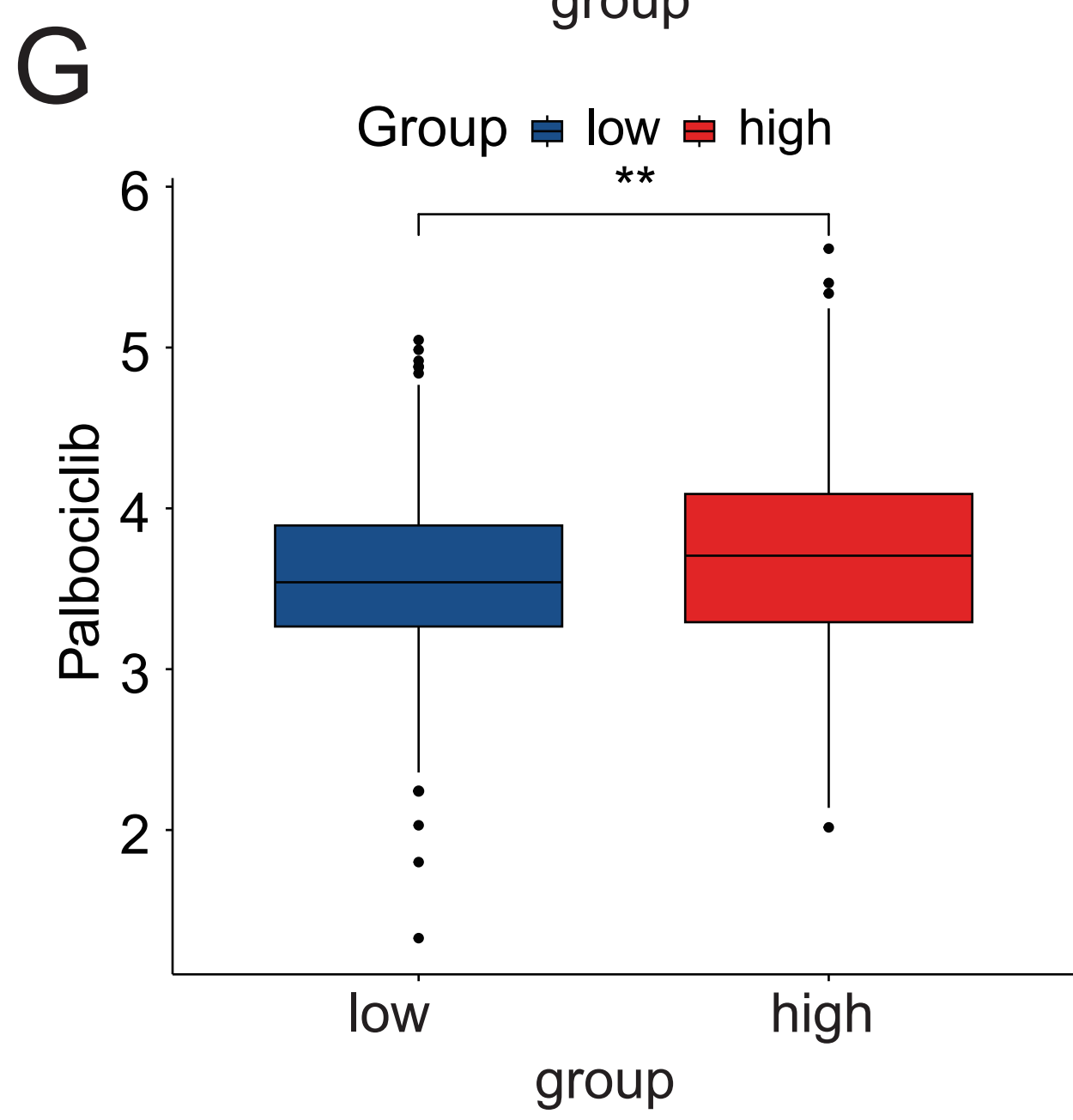
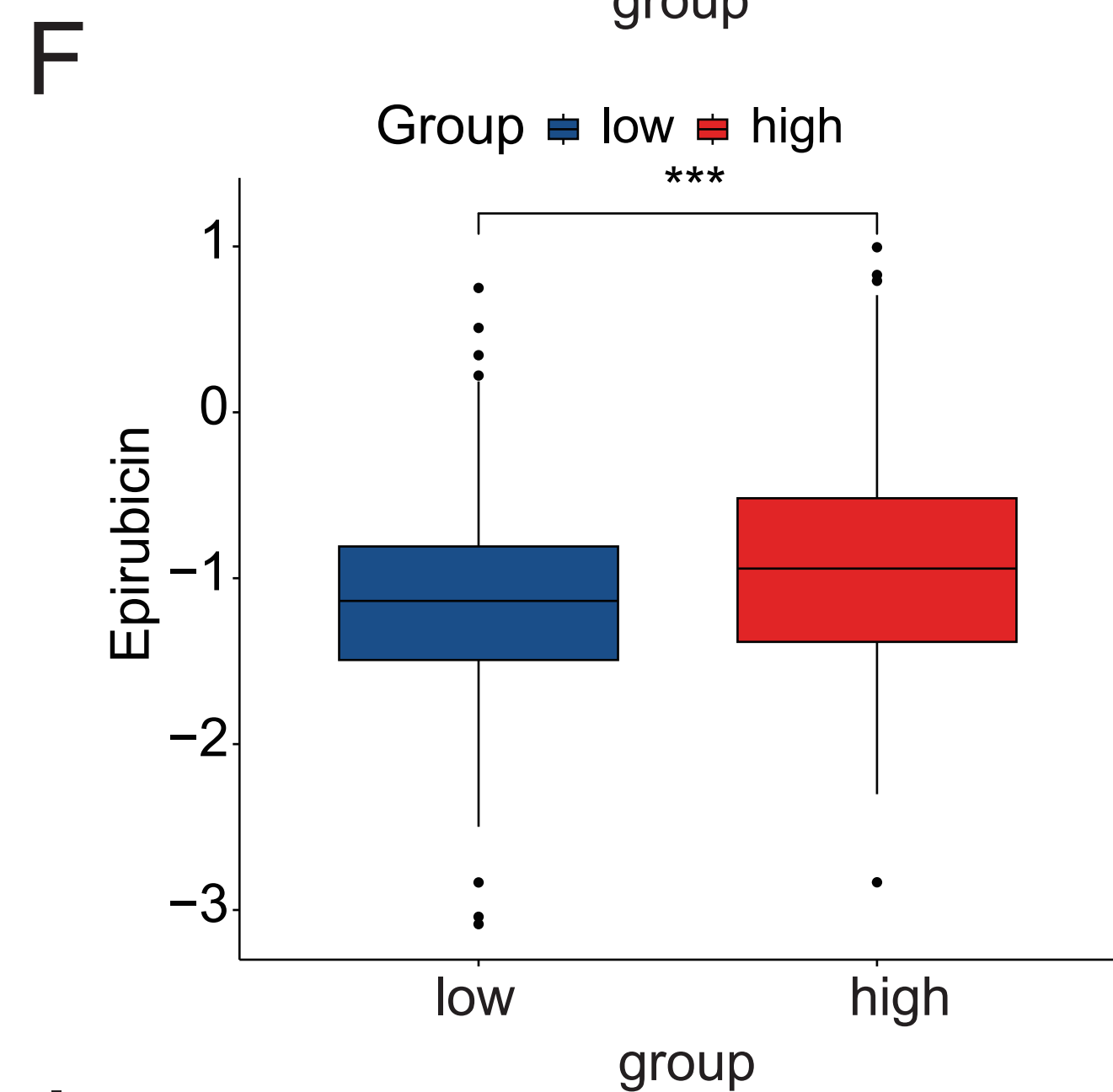
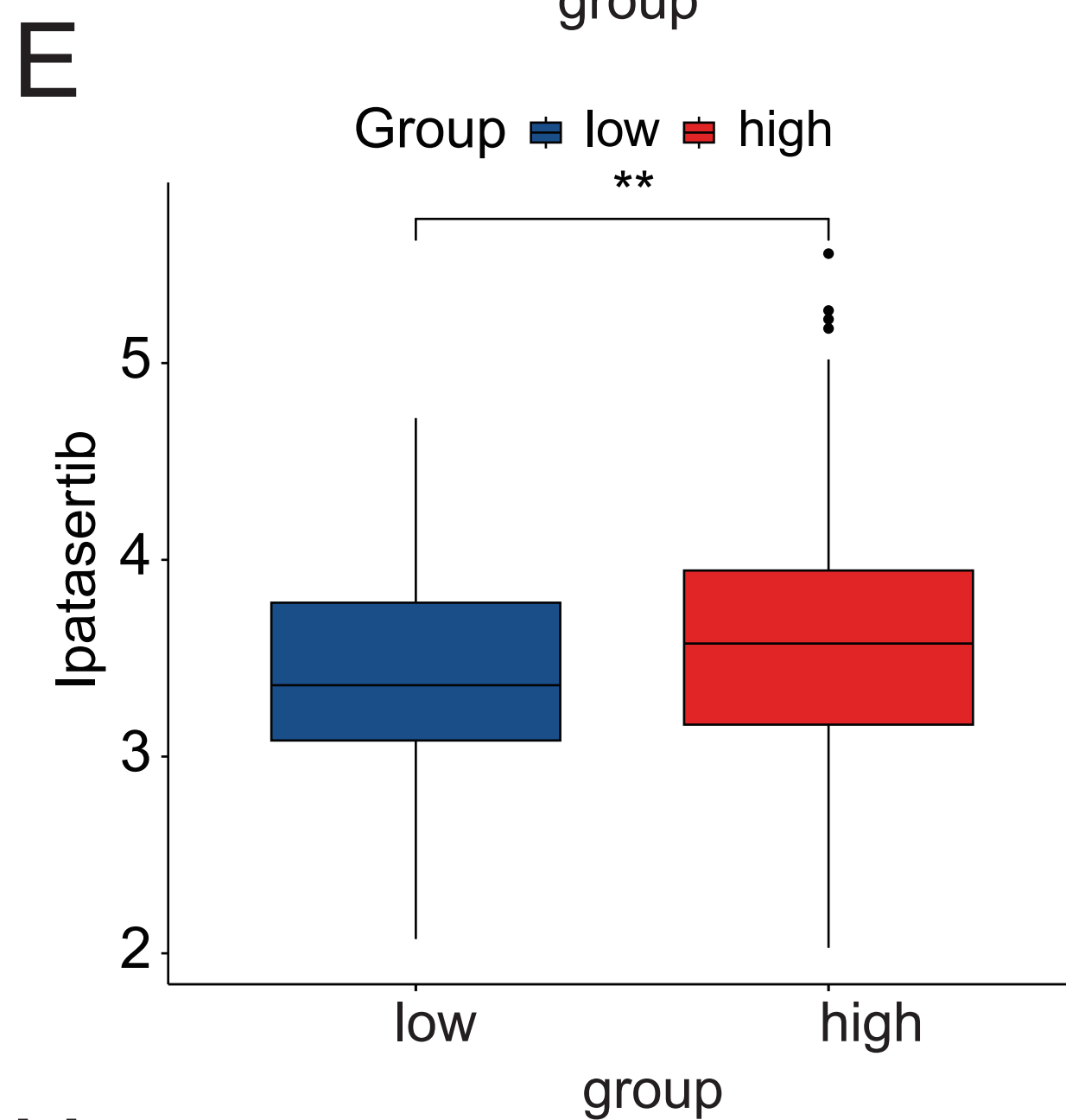
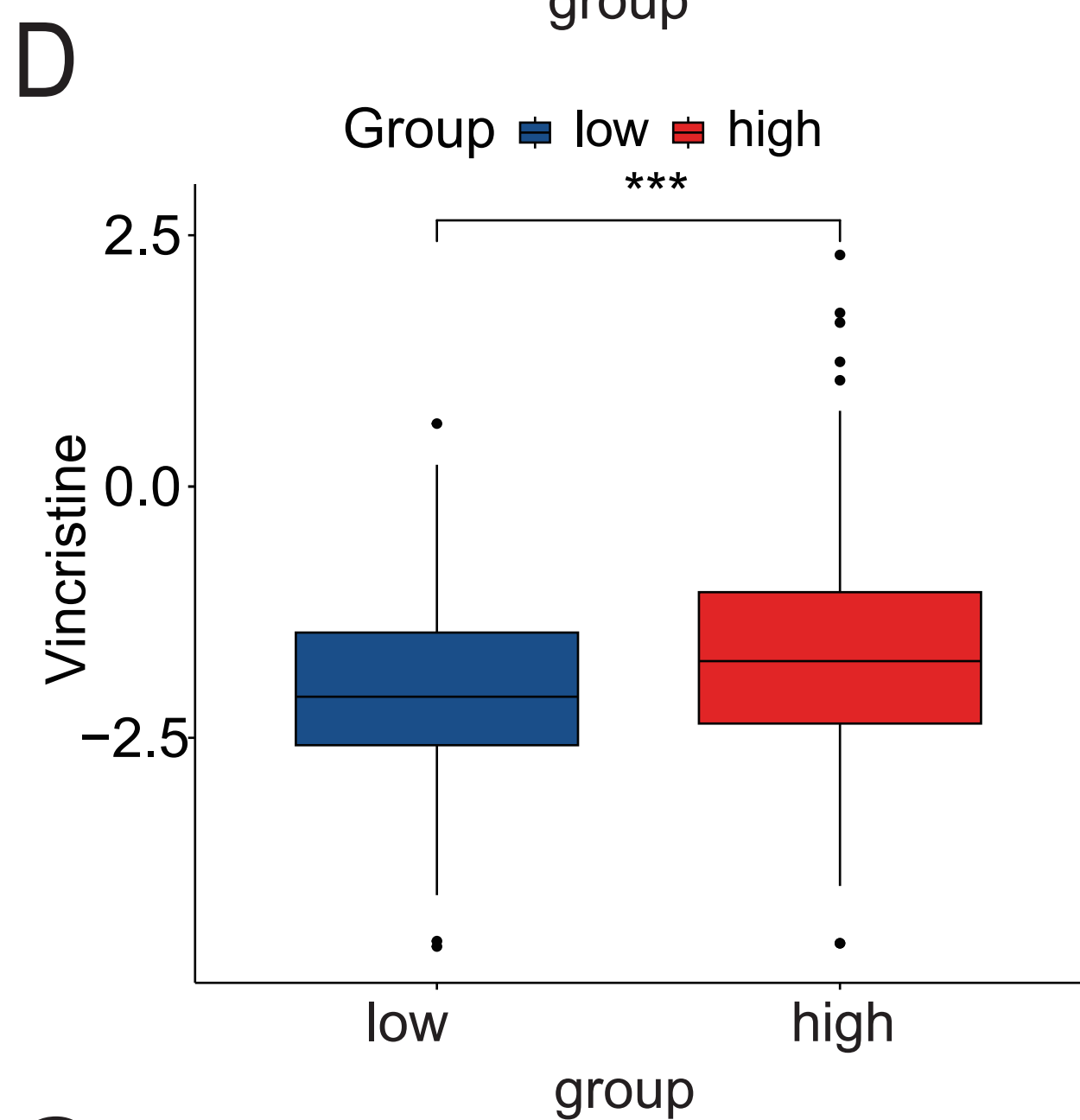
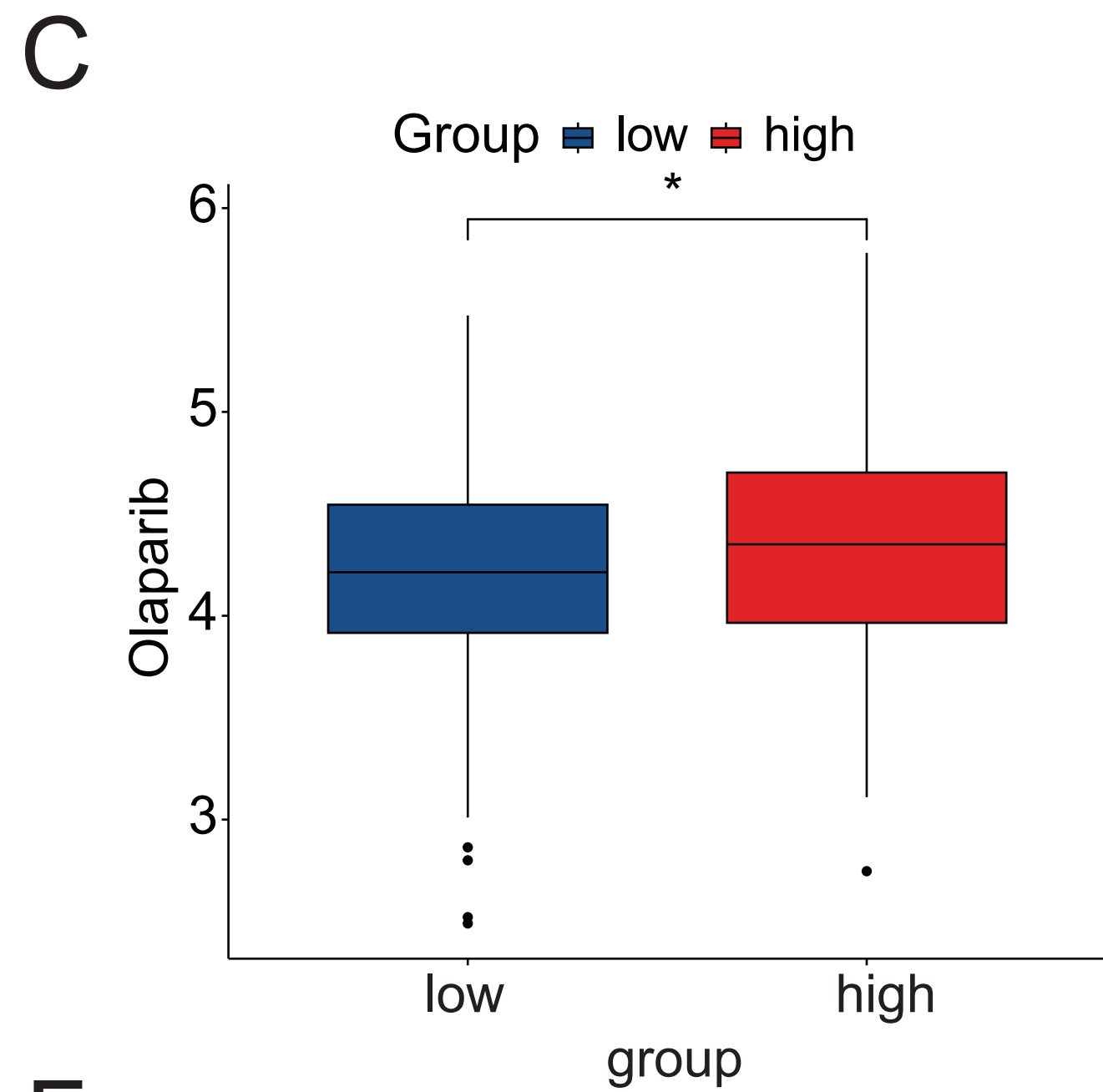
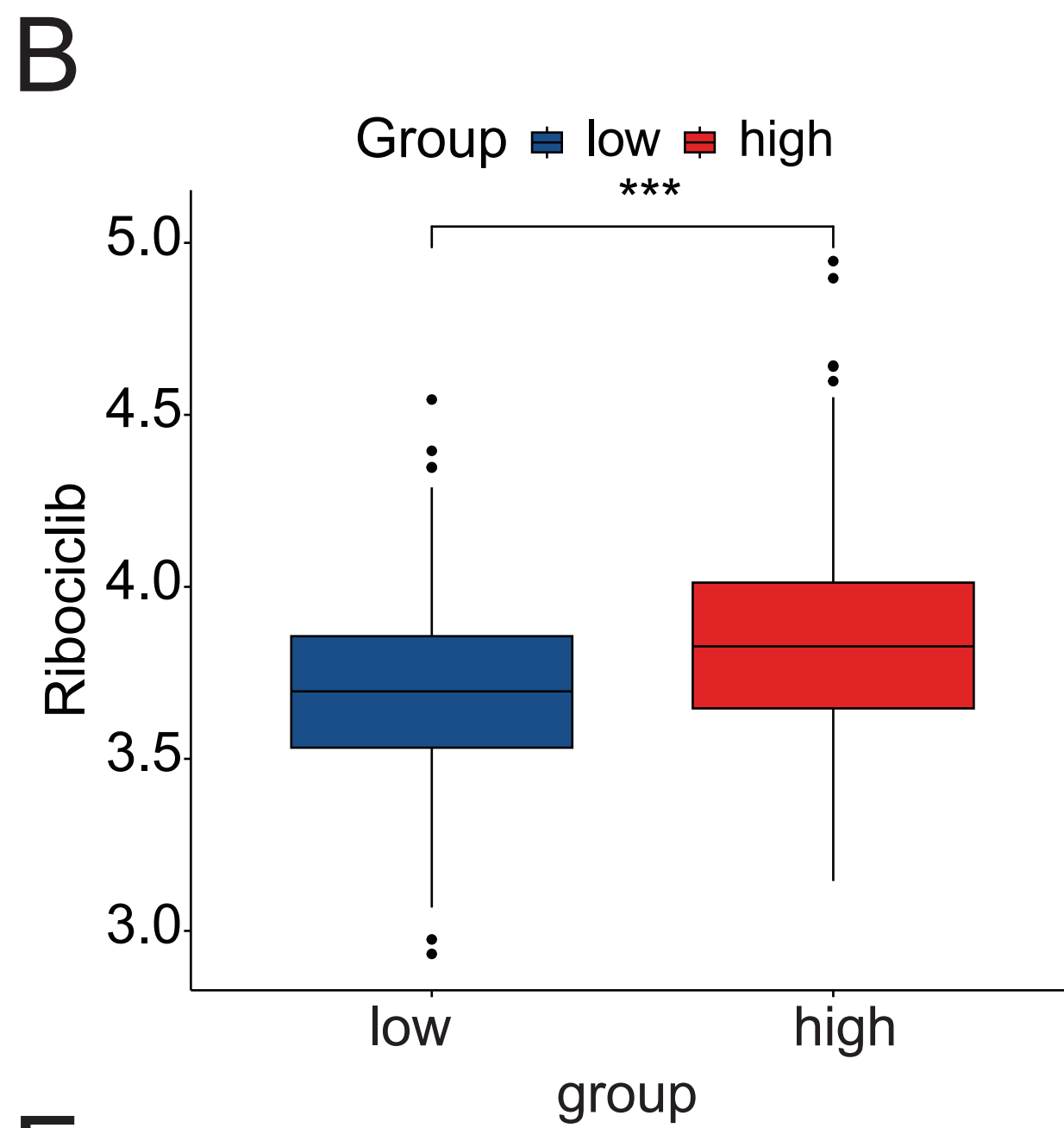
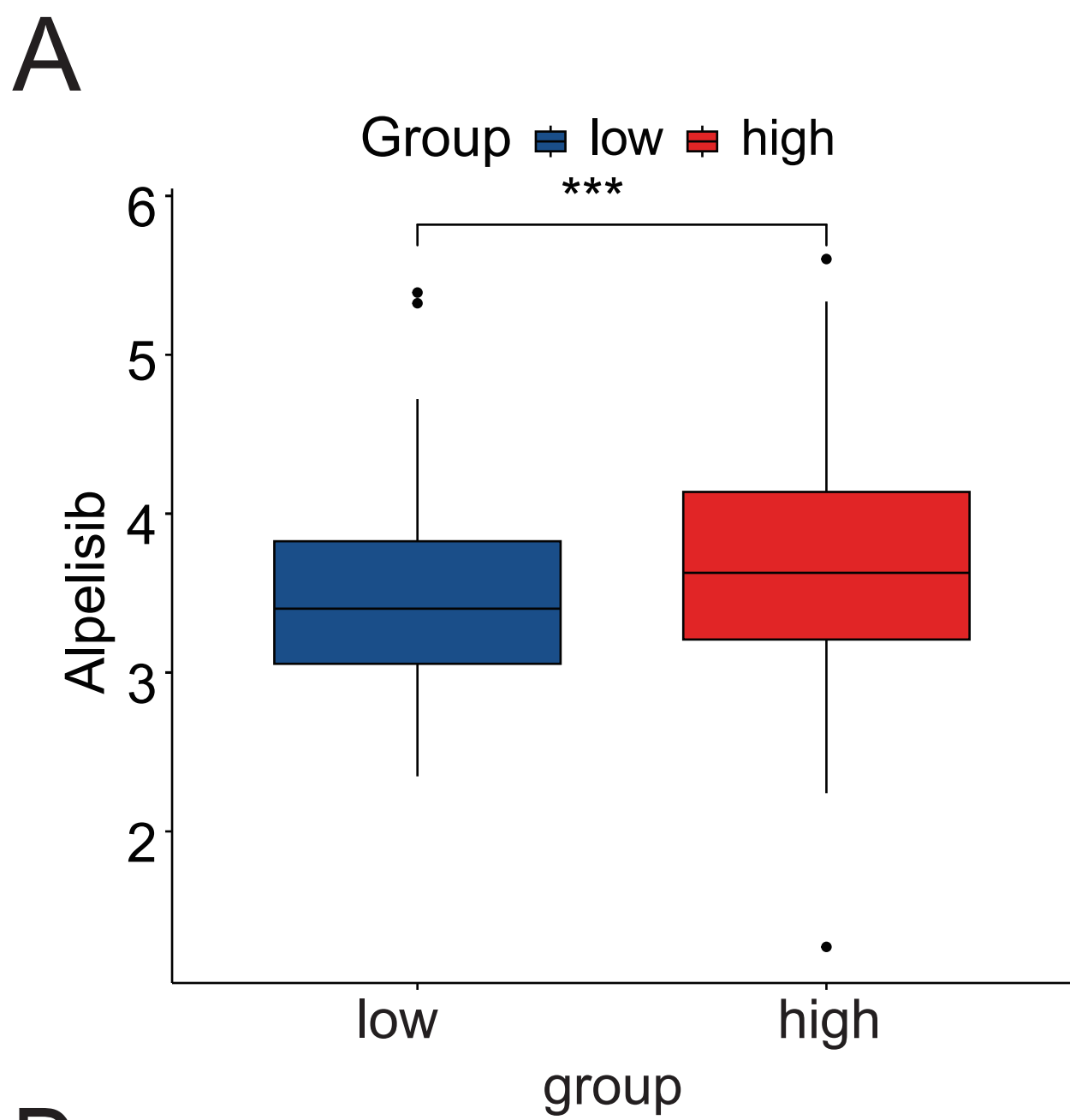


D

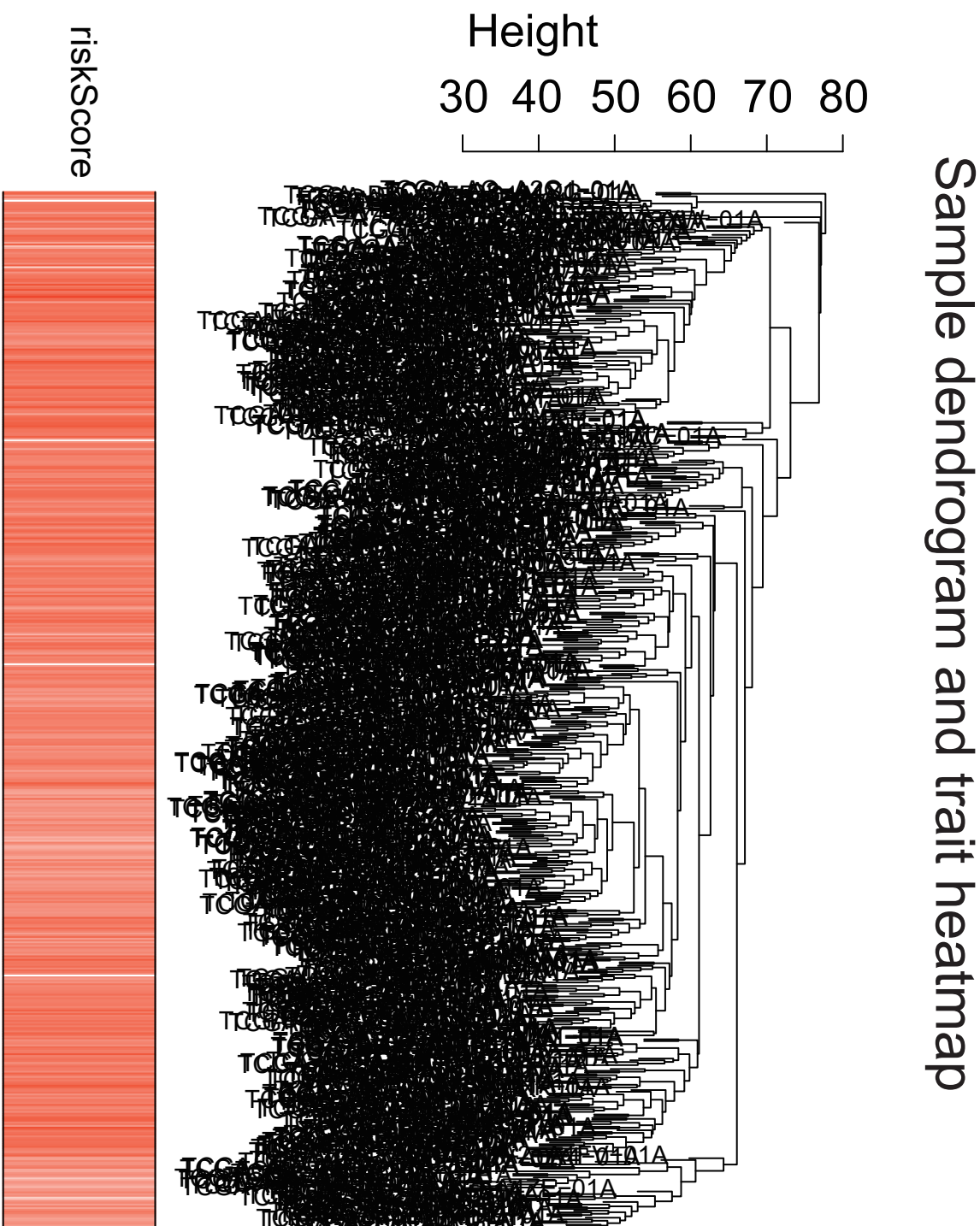


E

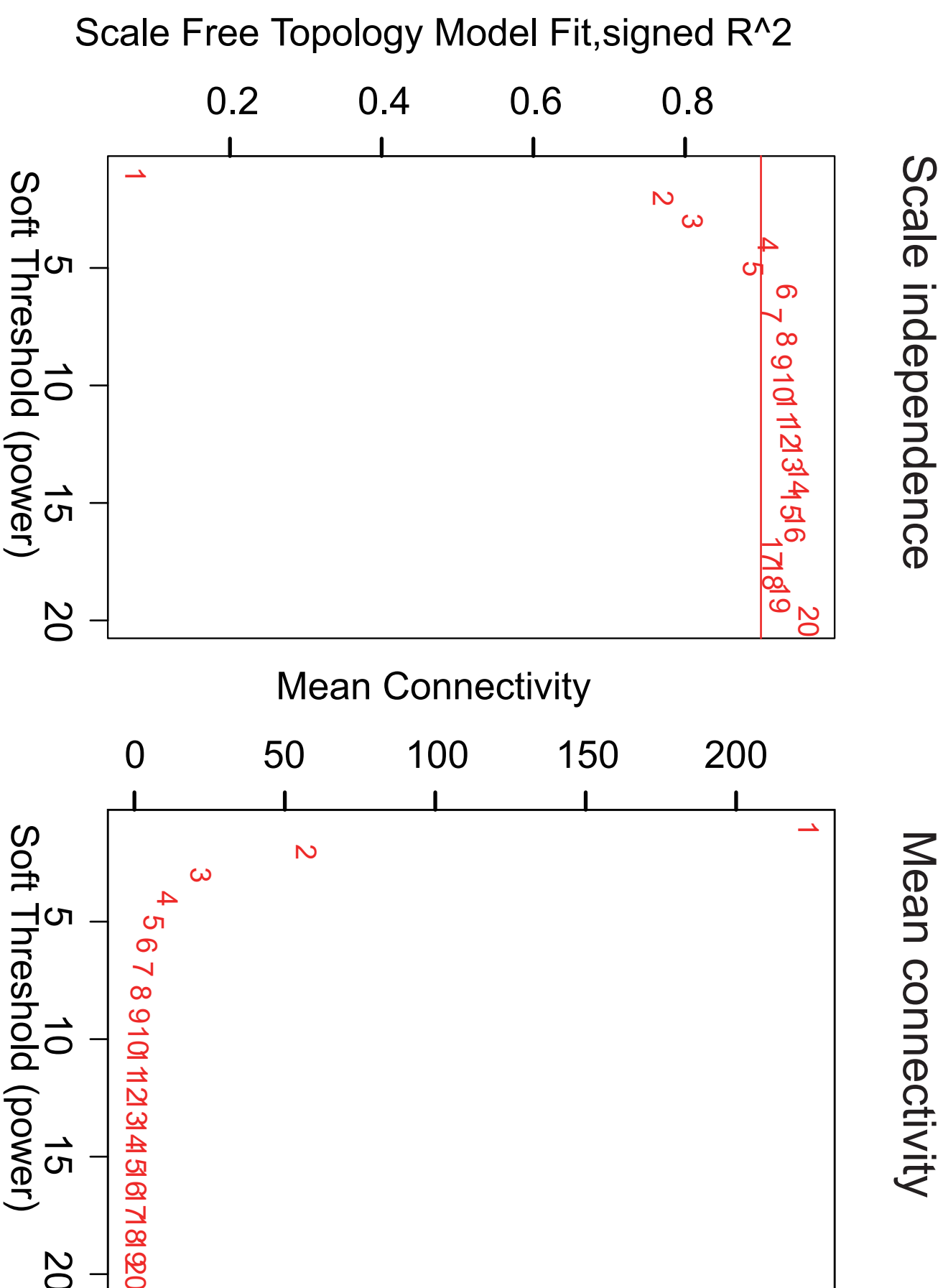




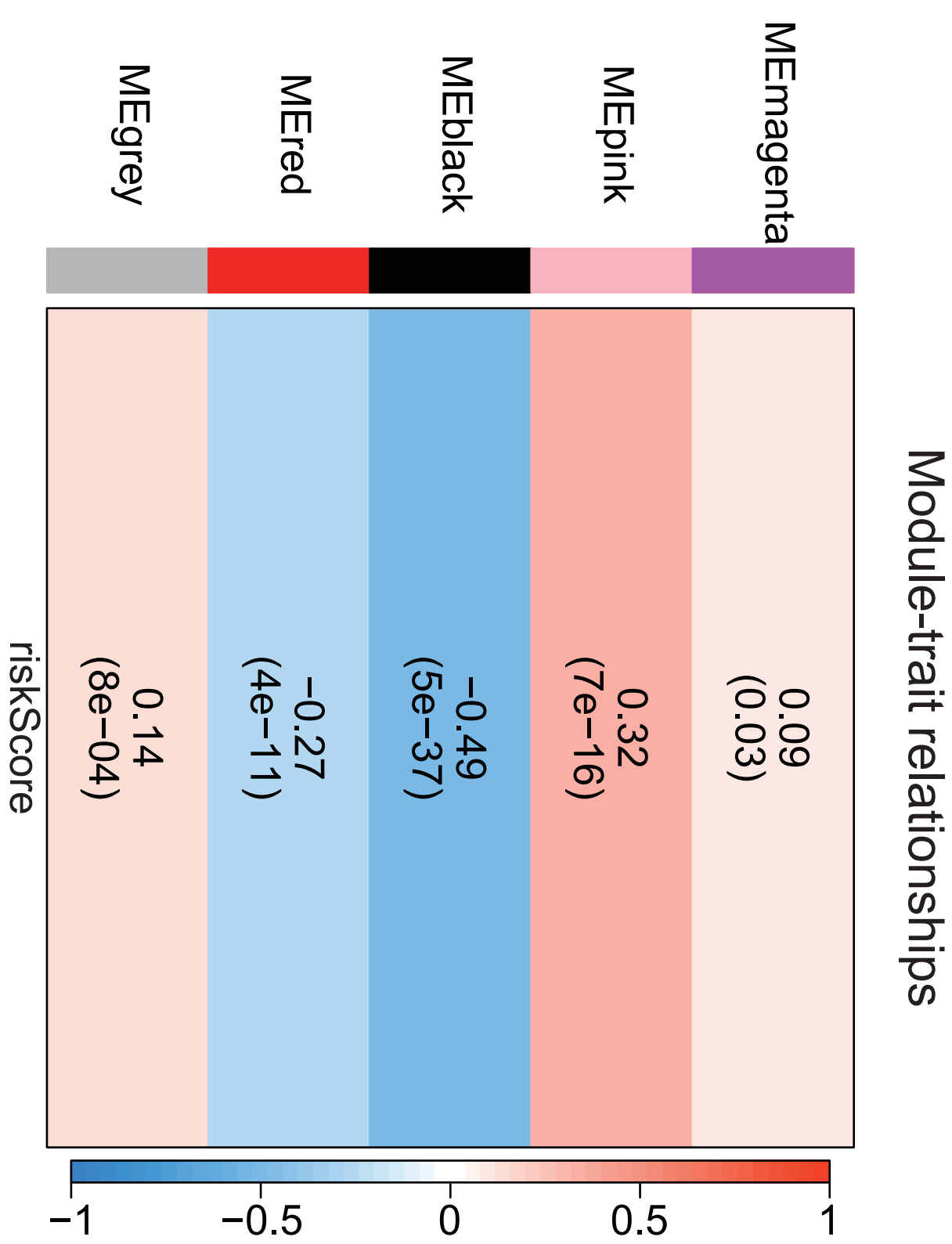
A



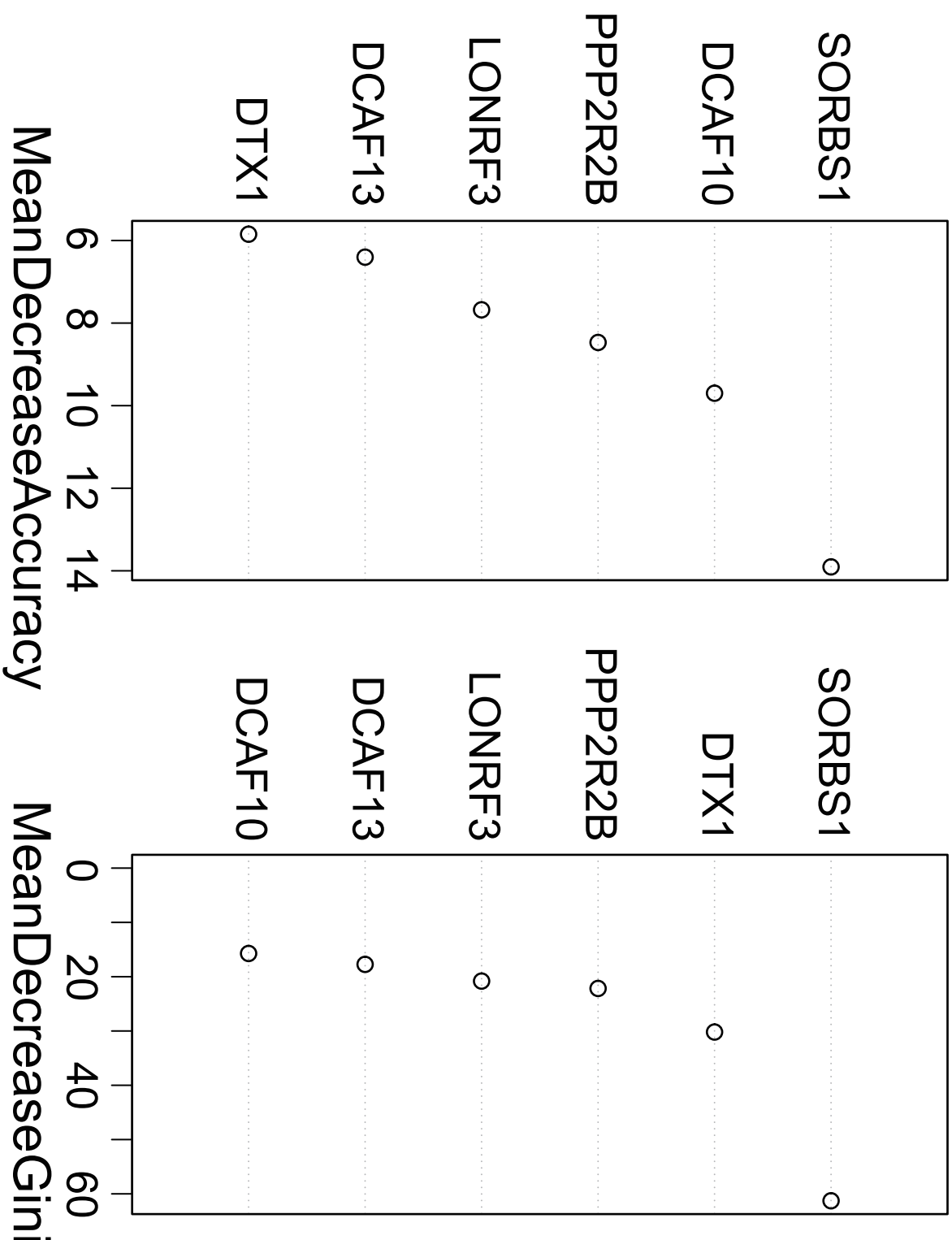
B



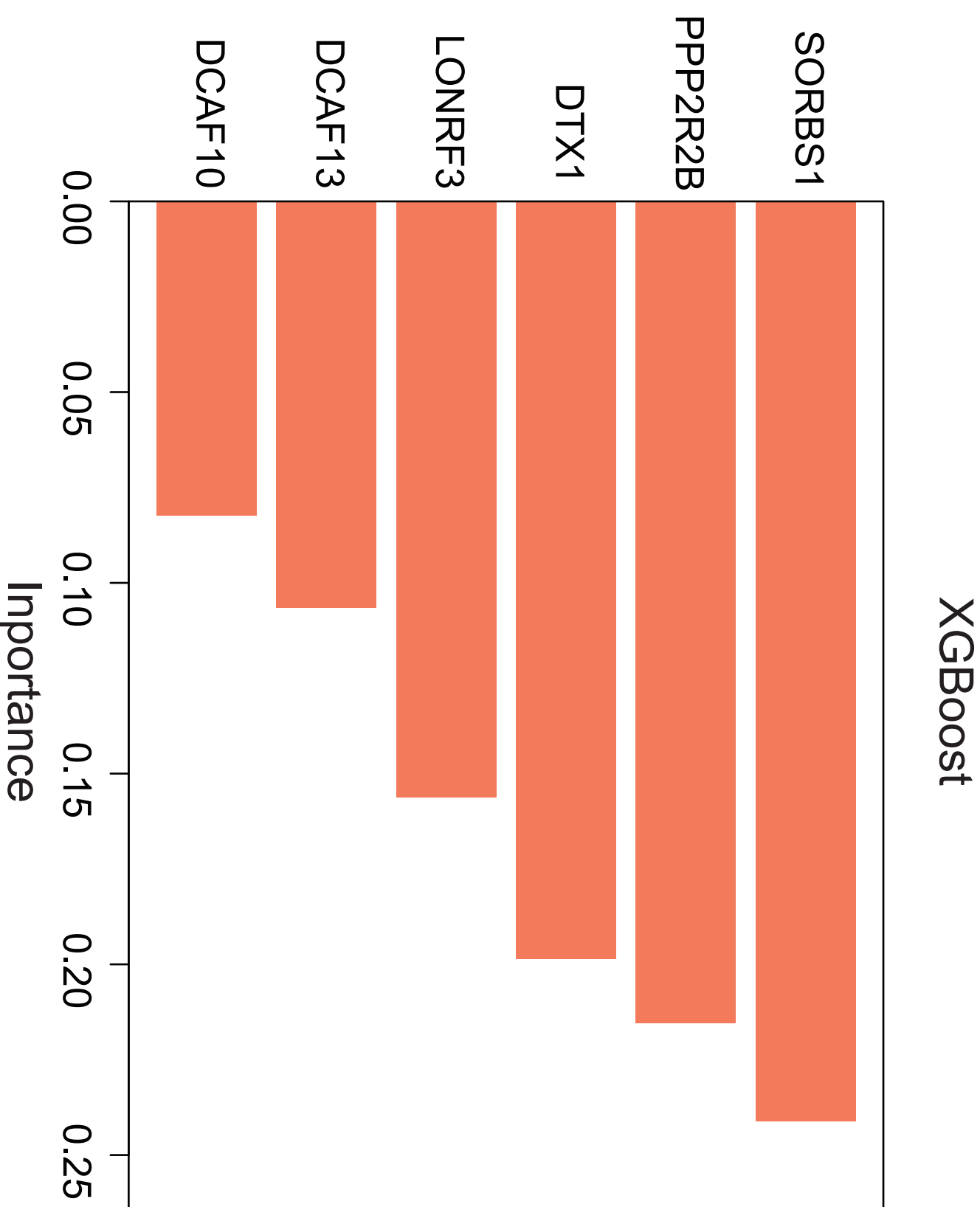
C



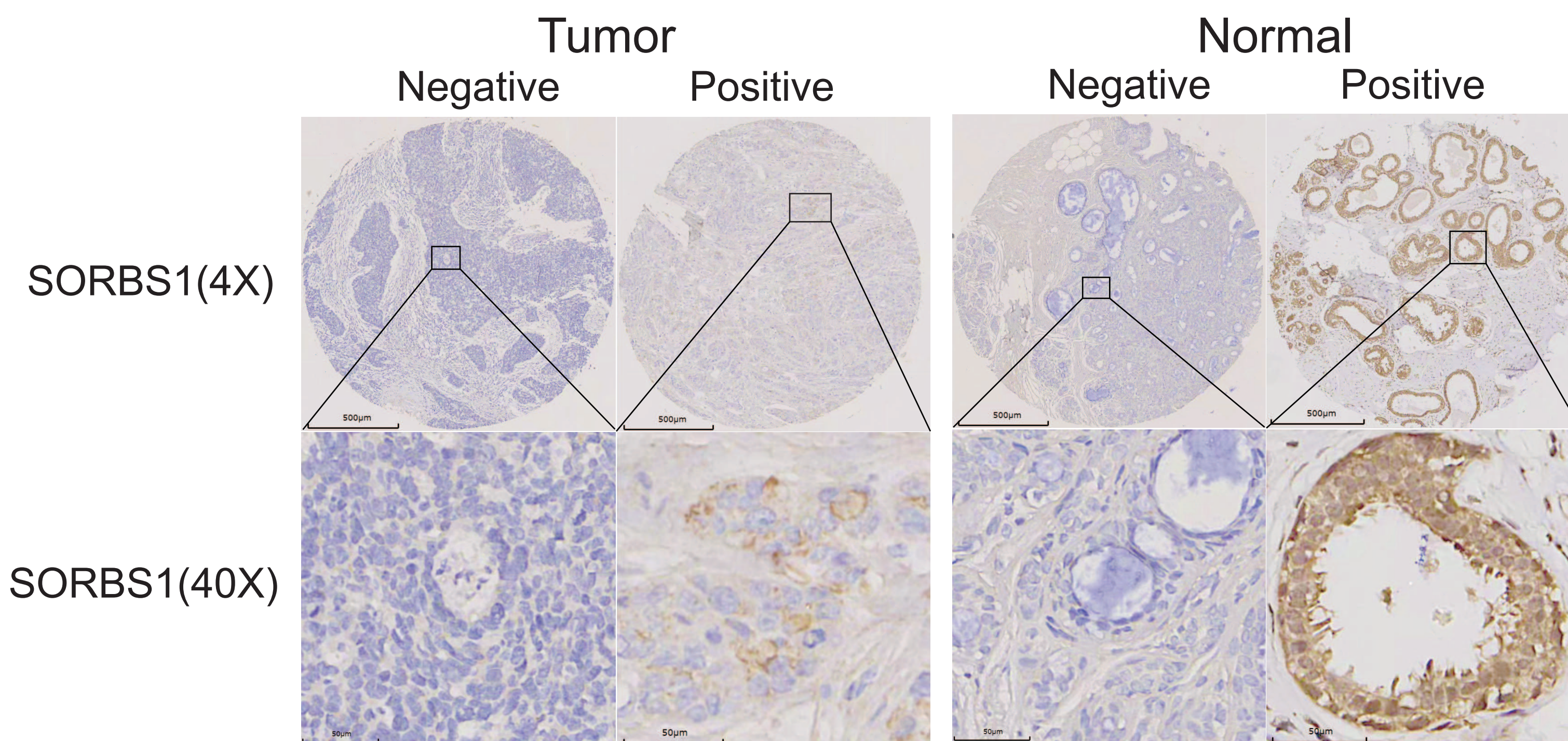
D



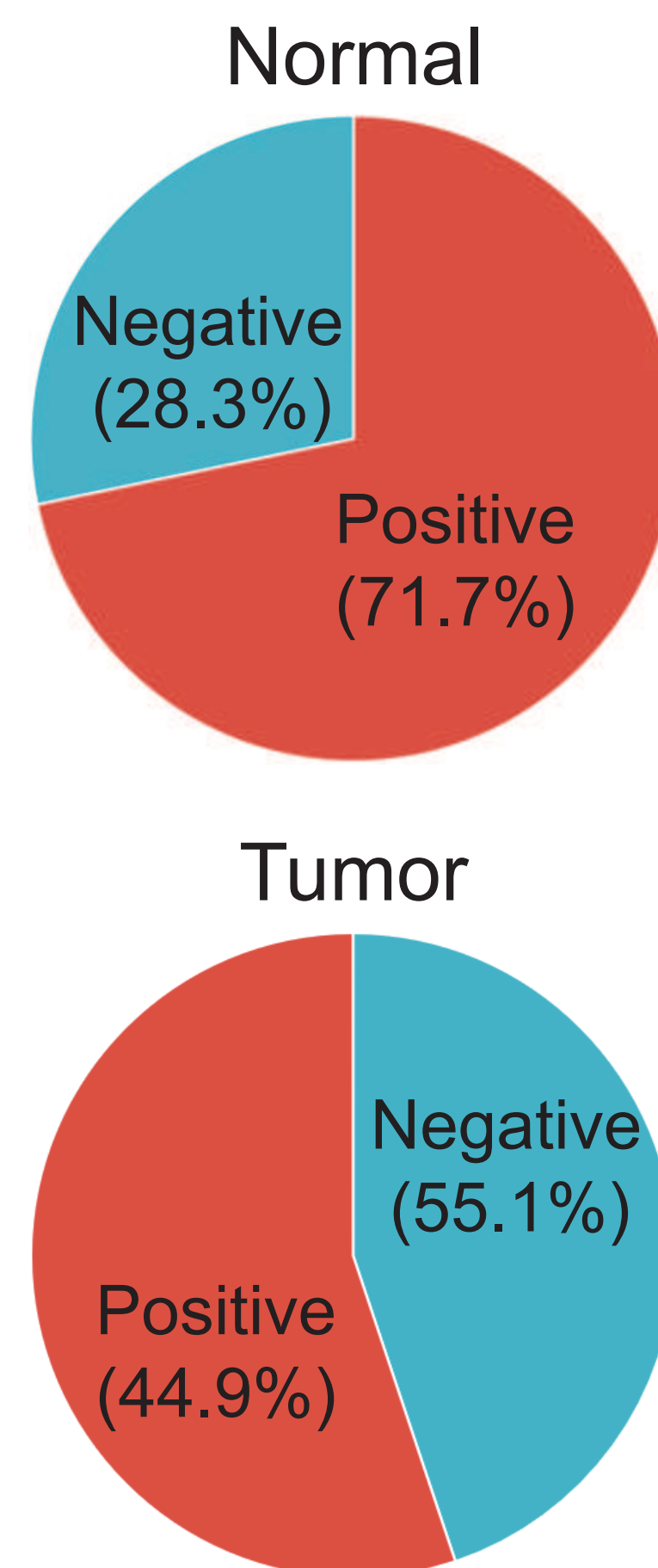
E



A



B



C

