Biomolecules and Biomedicine
ISSN: 2831-0896 (Print) | ISSN: 2831-090X (Online)
Journal Impact Factor® (2025): 2.8
CiteScore® (2025): 4.7
www.biomolbiomed.com | blog.bjbms.org

SUPPLEMENTAL DATA

Histopathology-guided multimodal pathomics identifies a *PLEKHA6*-positive epithelial state associated with phosphoinositide metabolism in esophageal cancer

Zhi Wang[1#], ZhenJie Li[1#], Nuo Yan[1#], WenXun Xiong[1], WeiQiang Feng[1], YunZhe Wang[1], Yiping Wei[1*], ShengKai Huang[2*]

[1]Department of Thoracic Surgery, The Second Affiliated Hospital, Jiangxi Medical College, Nanchang University, Nanchang, China;

[2]Department of Thoracic Surgery and Vascular Intervention, XinSteel Central Hospital, Xinyu, China.

*Correspondence to **ShengKai Huang**: hank2050@163.com.

[#]Equally contributed to this work: Zhi Wang, ZhenJie Li, and Nuo Yan.

Full article is available at the following link: Histopathology-guided multimodal pathomics identifies a *PLEKHA6*-positive epithelial state associated with phosphoinositide metabolism in esophageal cancer

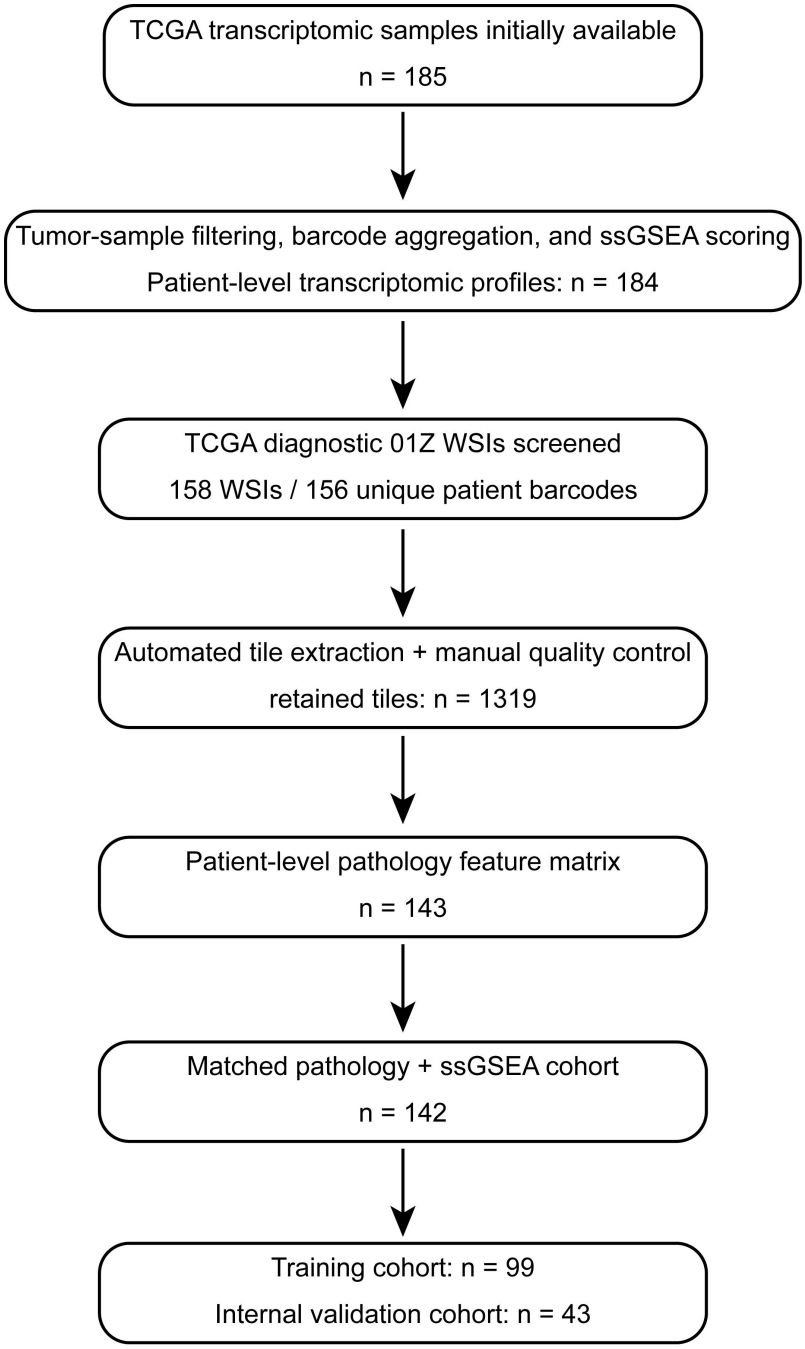


Figure S1. Patient and sample selection flowchart

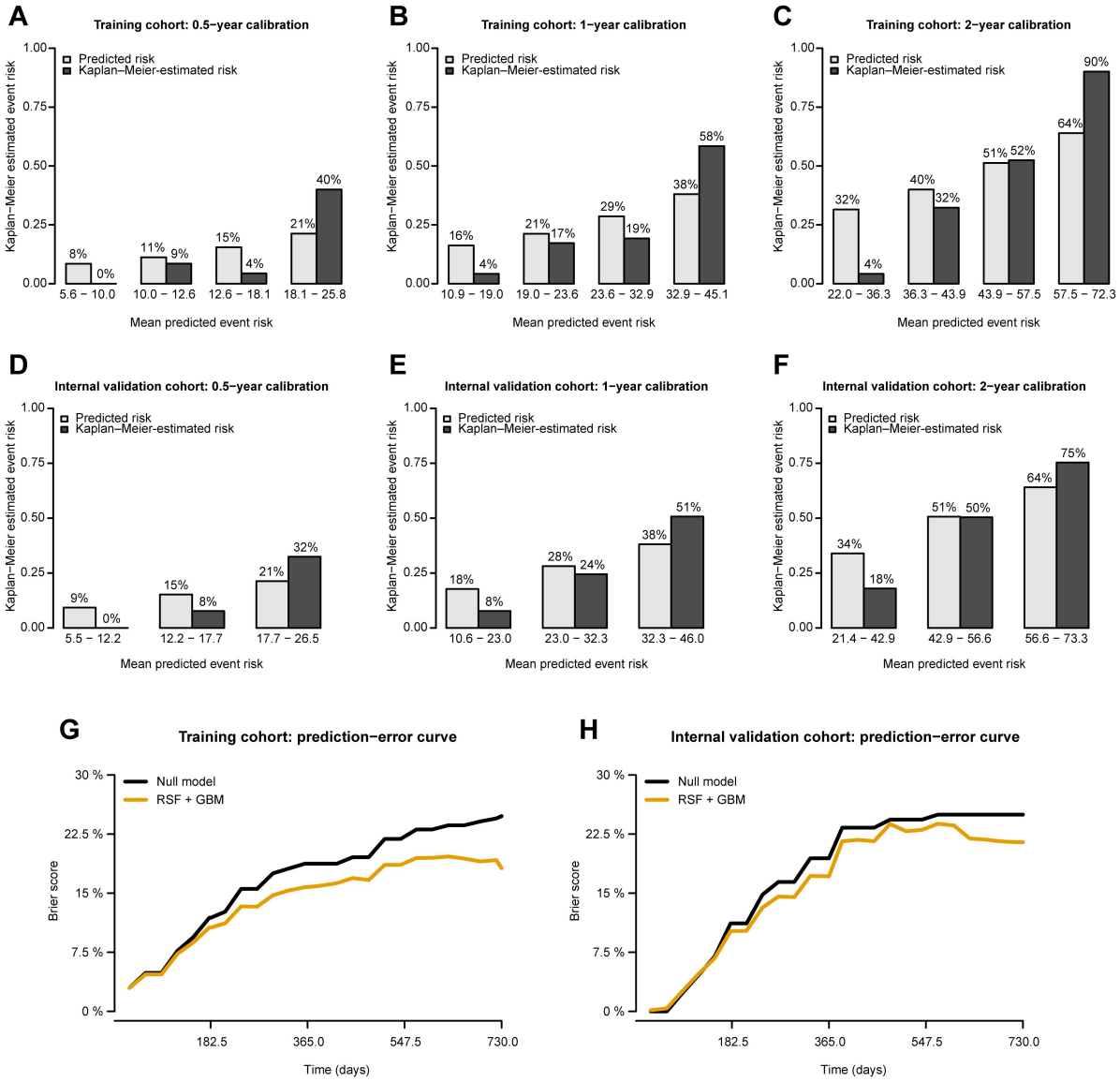


Figure S2. Exploratory calibration and prediction-error assessment of the frozen RSF plus GBM model. (A–C) Calibration at 0.5, 1, and 2 years in the training cohort using four quantile-based predicted-risk groups. **(D–F)** Corresponding calibration in the internal validation cohort using three predicted-risk groups. Within each group, mean predicted event risk was compared with Kaplan–Meier-estimated event risk; closer agreement between paired bars indicates better calibration. **(G, H)** Time-dependent IPCW Brier-score curves for the frozen model and the no-predictor null model in the training and internal validation cohorts, respectively. Lower Brier scores indicate lower prediction error. These analyses represent exploratory internal assessments and do not constitute independent validation. **Abbreviations:** GBM, gradient boosting machine; IPCW, inverse probability of censoring weighting; RSF, random survival forest.

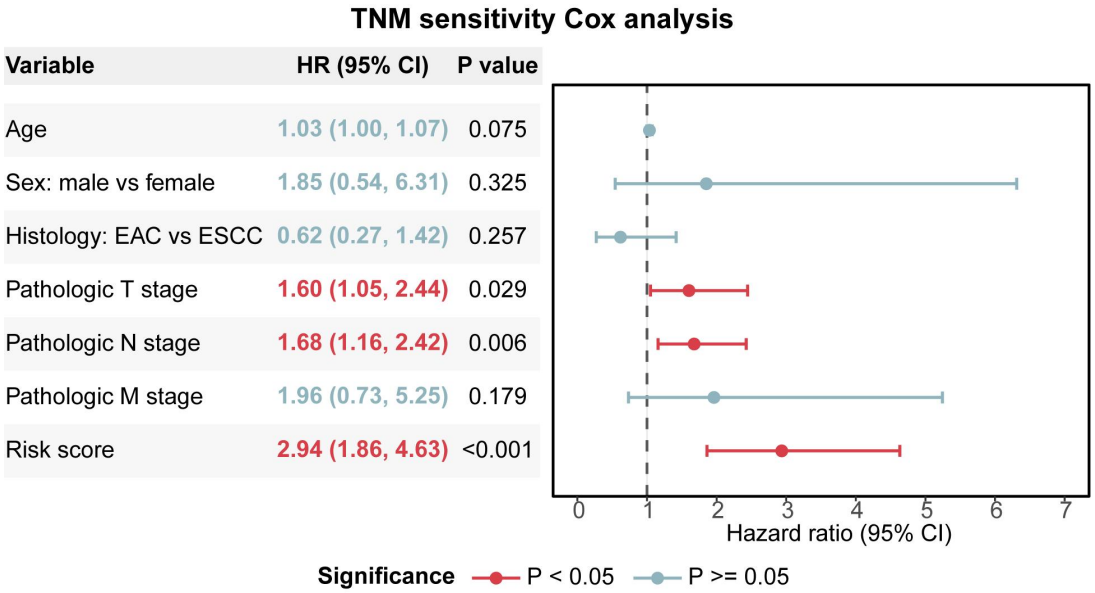


Figure S3. TNM-based sensitivity analysis of the association between the pathology-derived risk score and overall survival. Forest plot of the multivariable Cox model including the standardized pathology-derived risk score, age, sex, histological subtype, and pathological T, N, and M stages. The HR for the risk score represents a one-SD increase. Points indicate HRs, horizontal lines indicate 95% CIs, and red denotes $P < 0.05$. **Abbreviations:** CI, confidence interval; EAC, esophageal adenocarcinoma; ESCC, esophageal squamous cell carcinoma; HR, hazard ratio; OS, overall survival; SD, standard deviation; TNM, tumor–node–metastasis.

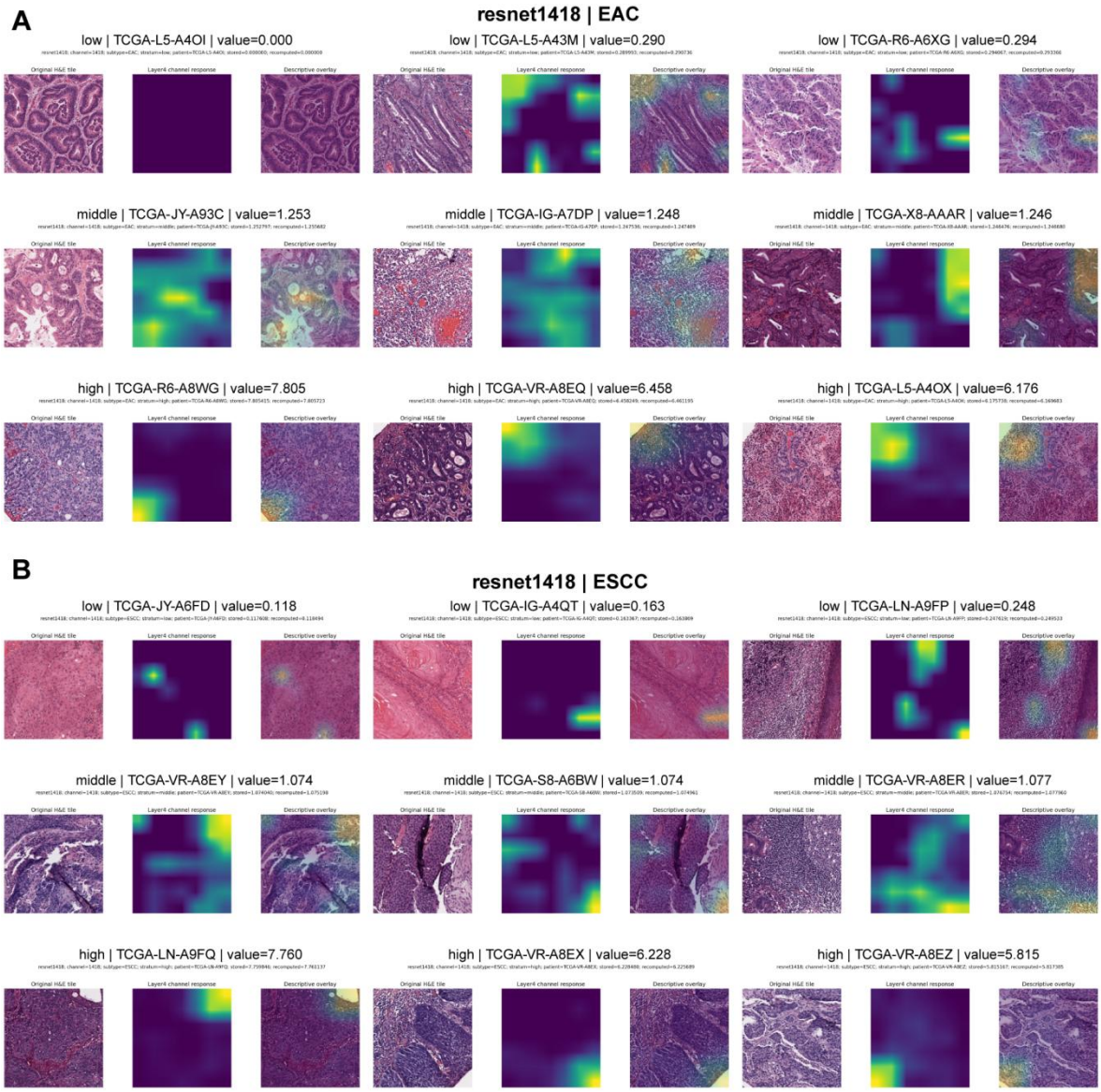


Figure S4. Channel-level spatial response maps for ResNet-50 feature resnet1418. **(A)** Representative tiles from esophageal adenocarcinoma (EAC). **(B)** Representative tiles from esophageal squamous cell carcinoma (ESCC). For each histological subtype, three tiles from different patients are shown for the low-, middle-, and high-response strata of the tile-level resnet1418 distribution. Each example includes the original H&E tile, the upsampled spatial response map of ResNet-50 layer4 channel 1418, and the corresponding overlay. The reported value indicates the tile-level response of resnet1418. Channel numbering follows zero-based indexing. The native layer4 response map had a spatial resolution of 7 × 7 and was upsampled for visualization only. These maps provide descriptive spatial localization of the selected channel before spatial maximum pooling and should not be interpreted as survival-specific Grad-CAM explanations or pathologist-validated morphological annotations.

Abbreviations: EAC, esophageal adenocarcinoma; ESCC, esophageal squamous cell carcinoma; H&E, hematoxylin and eosin.

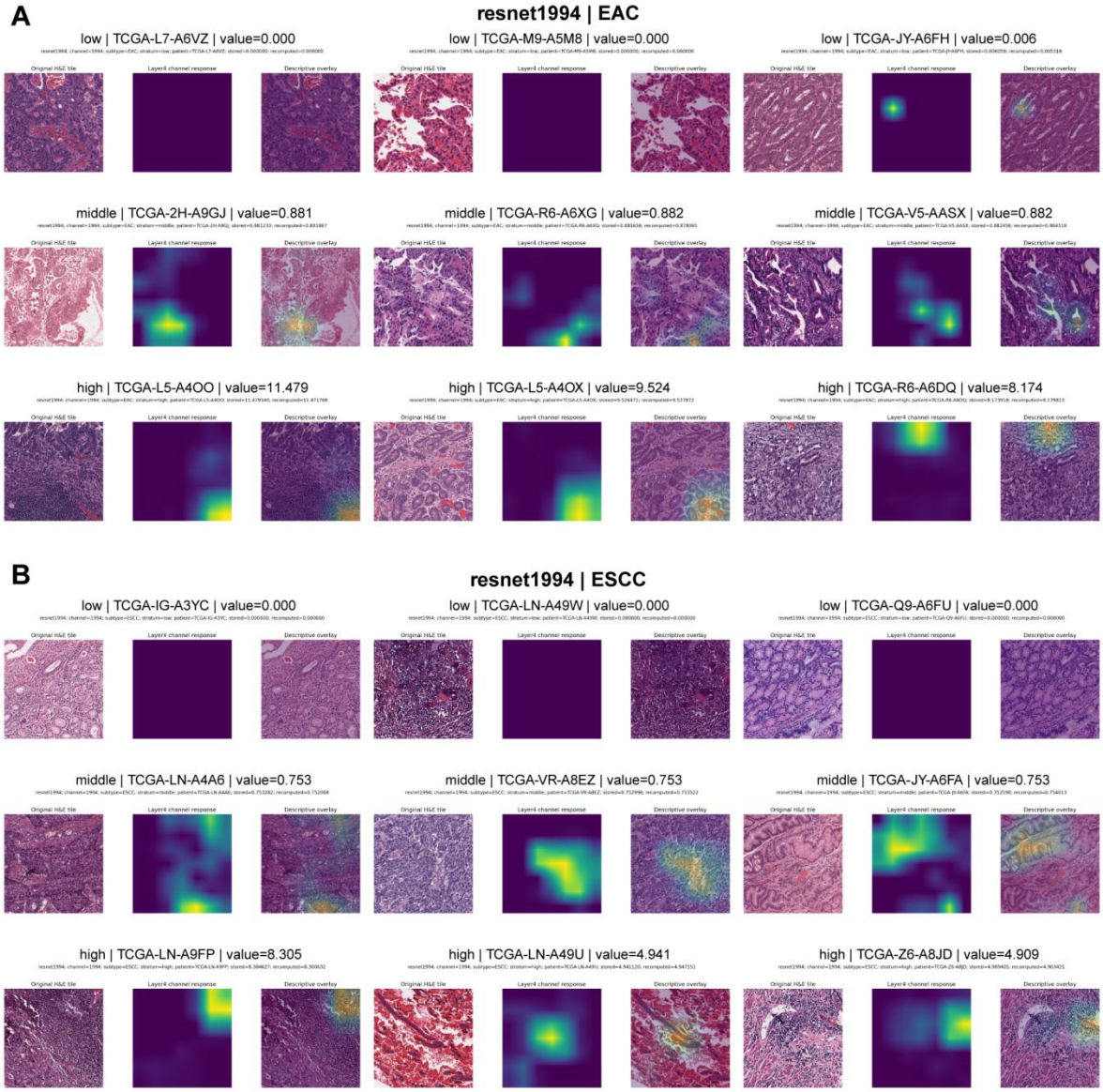


Figure S5. Channel-level spatial response maps for ResNet-50 feature resnet1994. **(A)** Representative tiles from esophageal adenocarcinoma. **(B)** Representative tiles from esophageal squamous cell carcinoma. For each histological subtype, three tiles from different patients are shown for the low-, middle-, and high-response strata of the tile-level resnet1994 distribution. Each example includes the original H&E tile, the upsampled spatial response map of ResNet-50 layer4 channel 1994, and the corresponding overlay. The reported value indicates the tile-level resnet1994 response. Channel numbering follows zero-based indexing. The native 7 × 7 response maps were upsampled for visualization only. These maps show descriptive channel responses before spatial maximum pooling and should not be interpreted as survival-

specific Grad-CAM explanations or pathologist-validated morphological annotations. **Abbreviations:** EAC, esophageal adenocarcinoma; ESCC, esophageal squamous cell carcinoma; H&E, hematoxylin and eosin.

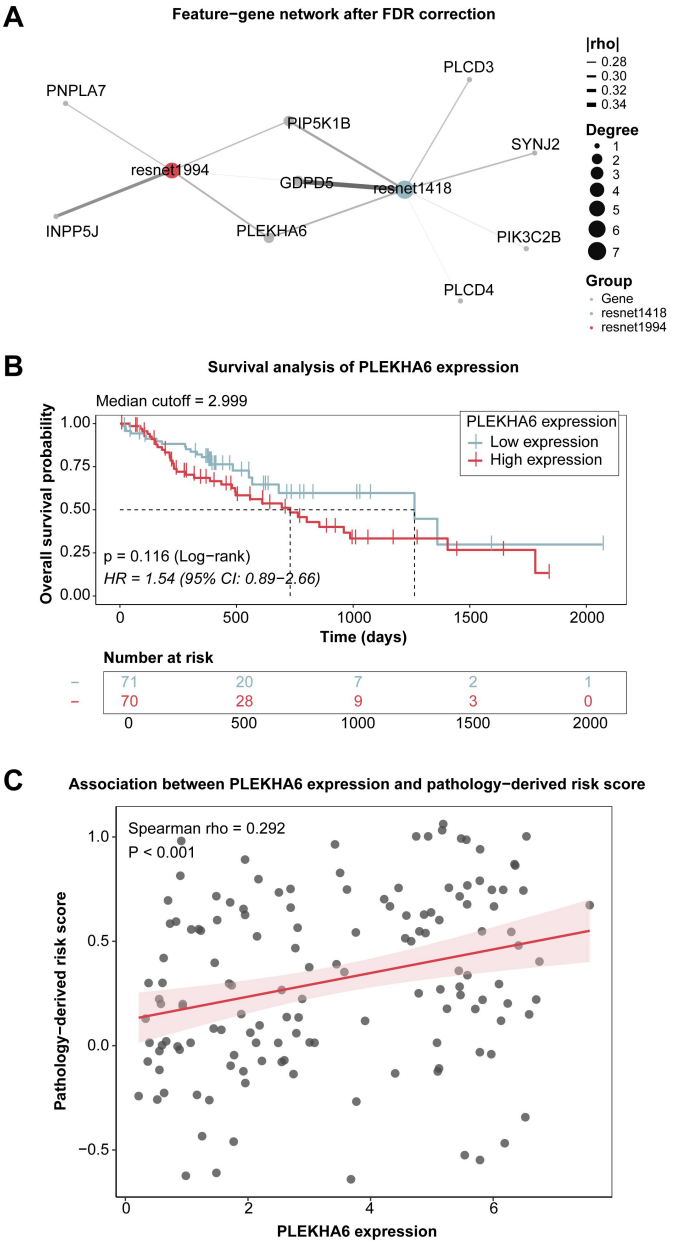


Figure S6. Sensitivity analyses supporting the prioritization of PLEKHA6 as a candidate pathology-associated gene. (A) Feature–gene correlation network after Benjamini–Hochberg correction. Edges represent Spearman correlations meeting FDR < 0.05 and $|\rho| \geq 0.15$; edge width indicates absolute correlation strength, and node size represents degree centrality. **(B)** Kaplan–Meier analysis of OS according to tumor-level *PLEKHA6* expression dichotomized at the median. Higher expression showed a directionally poorer survival pattern, but the difference was not statistically

significant. **(C)** Positive correlation between tumor-level *PLEKHA6* expression and the pathology-derived risk score, assessed using Spearman correlation analysis; the shaded area represents the 95% CI around the fitted regression line. **Abbreviations:** CI, confidence interval; FDR, false discovery rate; HR, hazard ratio; OS, overall survival.

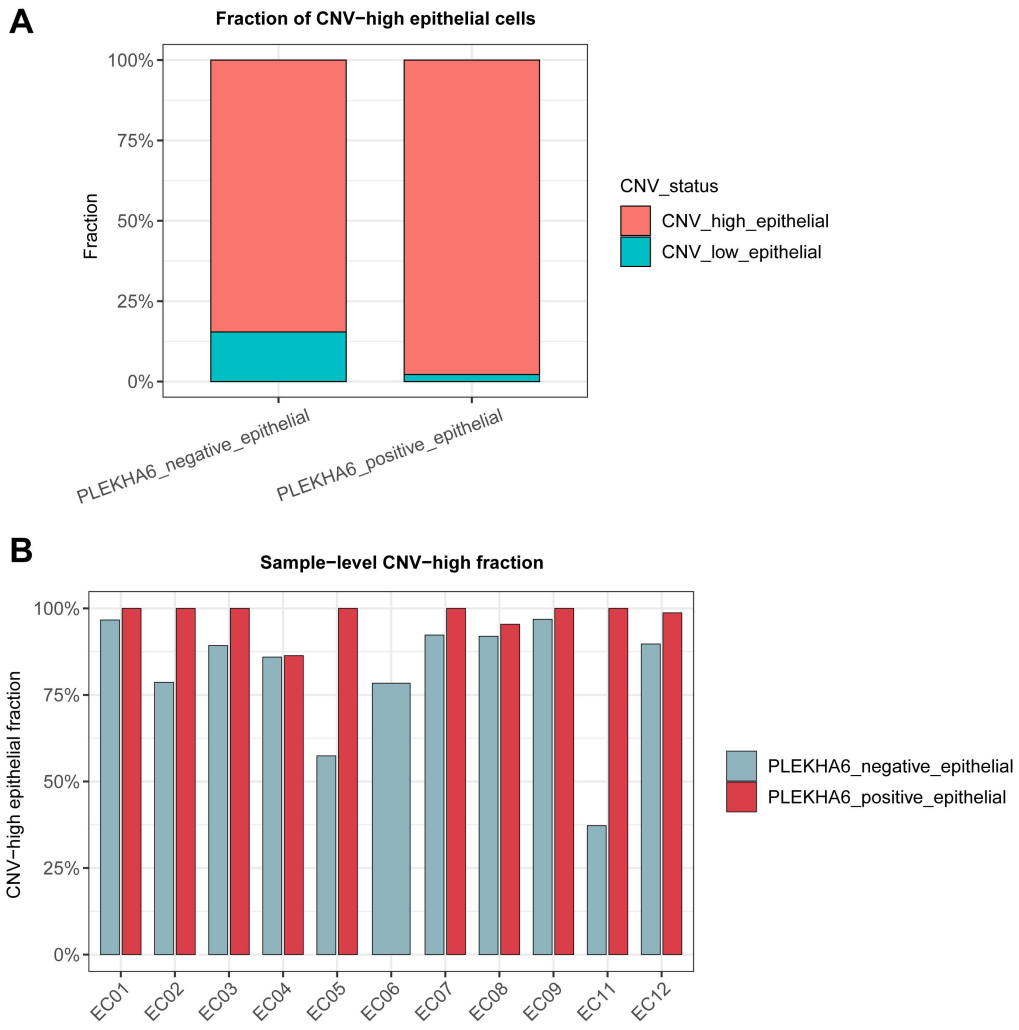


Figure S7. Exploratory inferCNV comparison of *PLEKHA6*-positive and *PLEKHA6*-negative epithelial cells. (A) Proportions of CNV-high and CNV-low cells in each epithelial group. CNV-high cells were operationally defined by inferCNV scores exceeding the 95th percentile of reference non-epithelial cells; the CNV-high fraction was greater in *PLEKHA6*-positive cells than in *PLEKHA6*-negative cells (97.80% vs 84.58%; Fisher's exact $P = 1.32 \times 10^{-33}$). **(B)** Sample-level CNV-high fractions across individual donors. The paired sensitivity analysis included eight samples containing both epithelial groups and at least five *PLEKHA6*-positive epithelial cells and showed a higher CNV-high fraction in the positive group (paired Wilcoxon $P = 0.0078$). Fractions from samples not meeting these criteria are shown

descriptively. inferCNV-derived signals indicate relative CNV-like transcriptional patterns and do not establish malignant-cell identity. **Abbreviations:** CNV, copy-number variation.

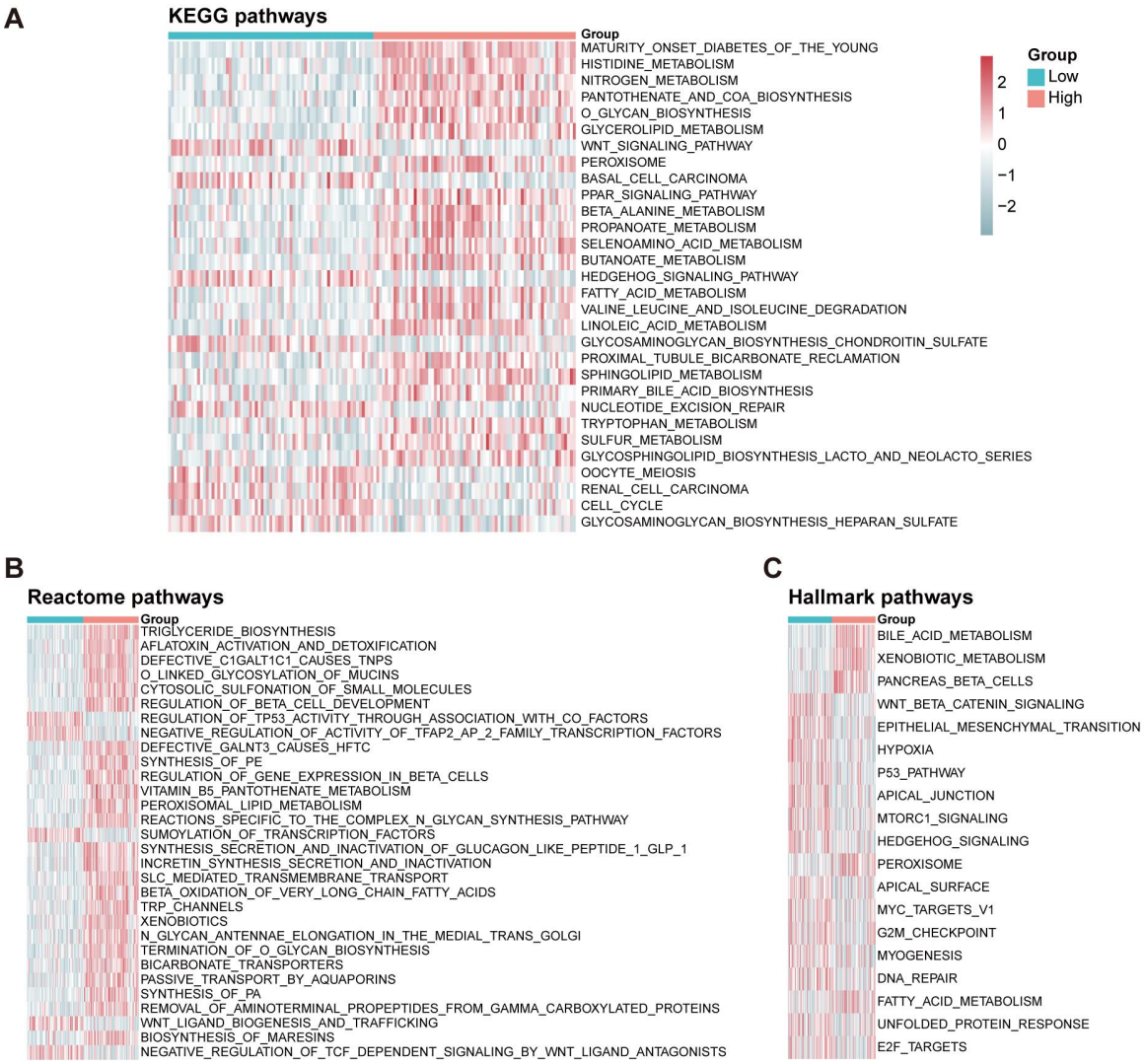


Figure S8. Bulk transcriptomic pathway profiles according to tumor-level PLEKHA6 expression. (A) KEGG pathway heatmap showing higher activity of multiple metabolic pathways in PLEKHA6-high tumors, including histidine, nitrogen, pantothenate/CoA, glycerolipid, fatty acid, linoleic acid, sphingolipid, tryptophan, and sulfur metabolism. **(B)** Reactome pathway heatmap showing enrichment of lipid metabolism, detoxification, glycosylation, vitamin B5, peroxisomal, and transport-related pathways in PLEKHA6-high tumors. **(C)** Hallmark pathway heatmap showing enrichment of bile acid metabolism, xenobiotic metabolism, pancreas beta cells, apical surface, apical junction, peroxisome, fatty acid metabolism, and unfolded protein response in PLEKHA6-high tumors, whereas PLEKHA6-low tumors showed

relatively higher activity of proliferation- and tumor progression-related programs, including MYC targets, G2M checkpoint, DNA repair, and E2F targets. Columns represent tumor samples grouped by PLEKHA6 expression, and rows represent pathways. **Abbreviations:** CoA, coenzyme A; KEGG, Kyoto Encyclopedia of Genes and Genomes.

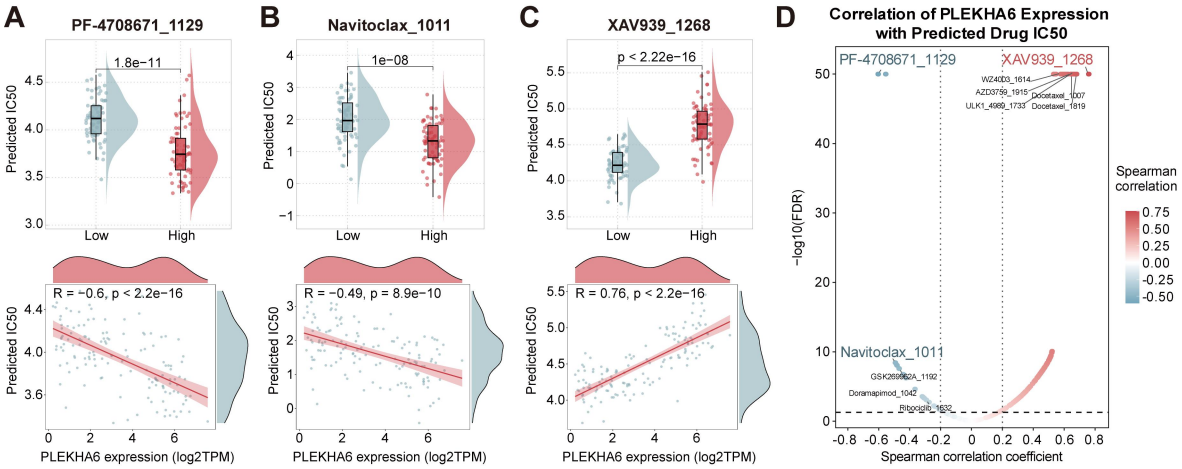


Figure S9. Exploratory predicted drug-response patterns associated with tumor-level PLEKHA6 expression. (A–C) Comparison of oncoPredict-derived IC50 values between PLEKHA6-low and PLEKHA6-high tumors for PF-4708671, navitoclax, and XAV939, respectively. PLEKHA6-high tumors showed lower predicted IC50 values for PF-4708671 and navitoclax, but higher predicted IC50 values for XAV939. Upper panels show group comparisons, and lower panels show the corresponding correlations with continuous PLEKHA6 expression. **(D)** Summary plot of Spearman correlations between continuous tumor-level PLEKHA6 expression and predicted IC50 values across all analyzed compounds, highlighting PF-4708671, navitoclax, and XAV939. These results represent transcriptome-based computational predictions and should be interpreted as hypothesis-generating rather than as evidence of actual therapeutic sensitivity. **Abbreviations:** FDR, false discovery rate; IC50, half-maximal inhibitory concentration.

Table S1. Phosphoinositide metabolism-related genes

Table S2. Performance of candidate models

Table S3. Sample-flow summary

Table S4. Tile quality-control records

Table S5. Clinicopathological characteristics of the cohort

Table S6. Histological-subtype and image-derived technical sensitivity analyses of pathology–PI metabolism associations

Table S7. Exploratory calibration and prediction-error assessment of the frozen RSF + GBM model

Table S8. Cox regression and proportional hazards tests

Table S9. Histological subtype sensitivity analyses

Table S10. ResNet–CellProfiler feature correlations

Table S11. Feature–gene correlations

Table S12. inferCNV analysis of epithelial cells

Table S13. Distribution of *PLEKHA6*-positive epithelial cells

Table S14. ESTIMATE and CIBERSORT analyses

Table S15. oncoPredict drug-response analysis

Supplemental file 1. CellProfiler project file for total-image feature measurement

Supplemental file 2. CellProfiler project file for object-recognition and object-level feature measurement