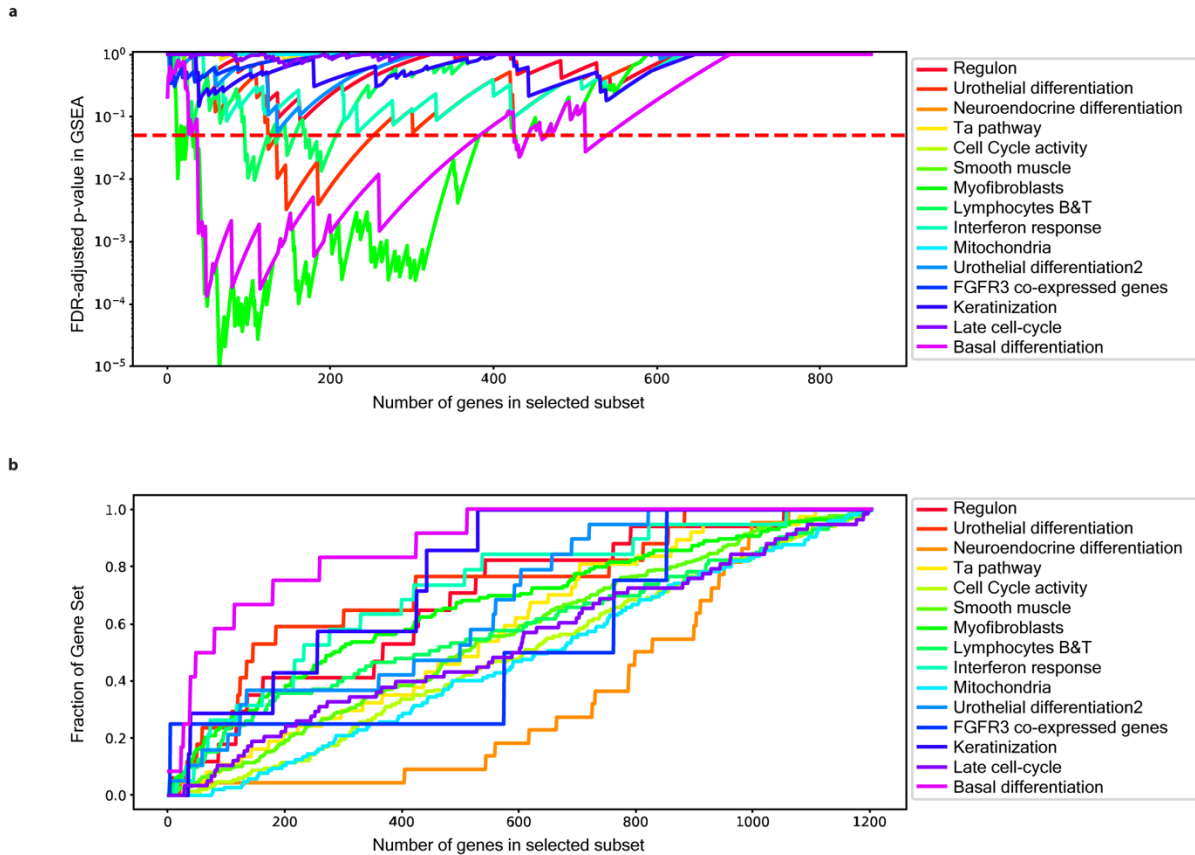


Predicting response to neoadjuvant chemotherapy in muscle-invasive bladder cancer via interpretable multimodal deep learning

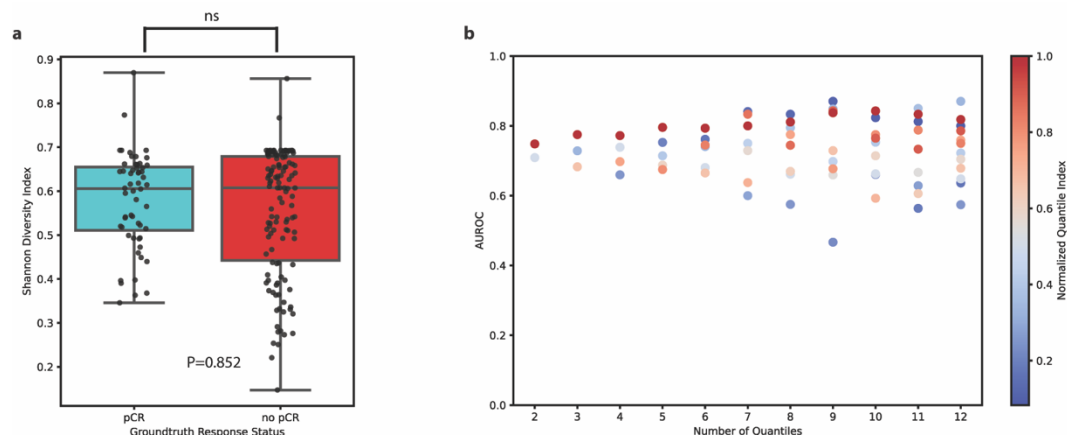
Supplementary Information

Supplementary Figures



Supplementary Figure 1. Sensitivity analysis for GSEA.

A sensitive analysis of GSEA was conducted using the top k gene aliases among the full gene alias list sorted by their average SHAP value magnitude in descending order. This analysis demonstrated the stability of identifying highly enriched gene sets among our study's 15 gene sets of interest. The x-axis represents k , the number of top gene aliases from the sorted list. **(a)** The y-axis shows the FDR-corrected hypergeometric test p-value, where the red dashed line denotes $P = 0.05$. **(b)** The x-axis shows the fraction of each of the 15 gene sets among the top k -selected genes.



Supplementary Figure 2. SDI-based ITH quantification stratified by response status and its influence on model performance.

The ITH quantification is computed with the Shannon Diversity Index (SDI) approach in (a) and (b). (a) Boxplots of ITH metrics from WSIs in pCR and no pCR subgroups. P-values computed by the Mann-Whitney U test. (b) Model performance evaluated by AUROC in different quantile subgroups stratified by ITH quantification. The x-axis indicates k, the number of quantiles, which ranges from 2 to the most significant number before the first appearance of invalid quantile subgroups for computing AUROC.

Supplementary Tables

Cohort Description		
Patient Characteristics	Categories	Number of Patients
Baseline Clinical T Stage	Clinical Stage T2	160
	Clinical Stage T3 or T4a	20
Baseline Zubrod Performance Status	Performance Status 0	137
	Performance Status 1	43
Gender	Male	150
	Female	30
Race	White	163
	Asian	4
	American Indian or Alaska Native	1
	Black or African American	4
	Unknown	8
Age	63.327±9.296 (Mean±Standard Deviation)	

Supplementary Table 1. Cohort Description of the SWOG S1314-COXEN clinical trial.

The clinical and demographic characteristics of the included patients are presented, with the number of patients stratified by clinical and demographic categories.

	5-fold CV NAC response	80 / 20 Testing NAC response
Patch Model (all matches)	0.5572±0.1793	0.5527
Patch Model (50% tumor purity)	0.5088±0.1373	0.4836
CLAM	0.4489±0.1451	0.6
Slidegraph+	0.6938±0.0565	0.6727

Supplementary Table 2. Unimodal model performance comparison using histological images to predict NAC response.

Performance comparison between three recently-developed weakly-supervised WSI-analysis approaches: 1. Patch-based model, 2. CLAM, and 3. SlideGraph+ using deep learning-derived features. The input to these models was the H&E-stained histology imaging data. The performance of each approach in predicting response to NAC was assessed by Area Under the Receiver Operating Characteristic Curve (AUROC). Both 5-fold cross-validation (5-fold CV) and testing performance on the hold-out test set were reported.

Supplementary Data List

Supplementary Data 1. SHAP-based Modality-level and Gene Importance Attribution.

The modality-level importance attributions for all patients in the hold-out test dataset were analyzed using a SHAP-based interpretation approach applied to a modality-level proxy model. For gene importance attribution analysis, SHAP was applied to a proxy model incorporating the gene expression feature vector alongside predictions from the two Graph Neural Network (GNN) branches.

Supplementary Filename: Supplementary_Data_1.xlsx

Index_Map: sheet index/number to table name mapping.

Supplementary Data 2. Patch-level Activations on WSI-derived Graphs.

The activation values of nodes in the graph are derived from each WSI and used as input to a GNN branch. Each node represents a specific patch or region of the WSI. A node value is assigned by a GNN branch to each node to measure its importance in the predictive model. The results for both the cell type and morphological features (denoted as “CTandMF” in the filename) GNN branch and the neural embeddings (denoted as “NE” in the filename) GNN branch are included.

Supplementary Filename: Supplementary_Data_2.xlsx

Index_Map: sheet index/number to table name mapping.

Supplementary Data 3. Quantification of cell-type-specific characteristics for *no-pCR* and *pCR* predictive patches, respectively, on each slide.

The average patch-level type-specific cell counts or tumor-stromal ratios for the most predictive patches for *no pCR* or *pCR* are normalized by the average patch-level cell-type-specific attribute of the entire WSI. The results for each distinct WSI are included. The patches or regions from the WSI with the top 25% (lower node values) and bottom 25% (higher node values) are extracted, which correspond to the most important patches being predictive for *pCR* and *no pCR*, respectively. The cell-type specific characteristics on each patch are quantified: the cell counts of cancer cells, connective cells, immune cells, and necrotic cells, as well as the tumor-stromal ratio calculated by dividing the cancer cell count by the connective cell count.

Supplementary Filename: Supplementary_Data_3.xlsx

Index_Map: sheet index/number to table name mapping.

Supplementary Data 4. Cell-type-specific cell counts in regions identified as predictive of different response statuses.

Cell counts of the different cell types - i.e., cancer, immune, necrotic, connective, and non-cancer, were analyzed for responder-associated and non-responder-associated regions. Patches enriched in specific cell types, particularly cancer cells, and associated with non-response or response were selected to identify potential histological markers.

Supplementary Filename: Supplementary_Data_4.xlsx

Index_Map: sheet index/number to table name mapping.

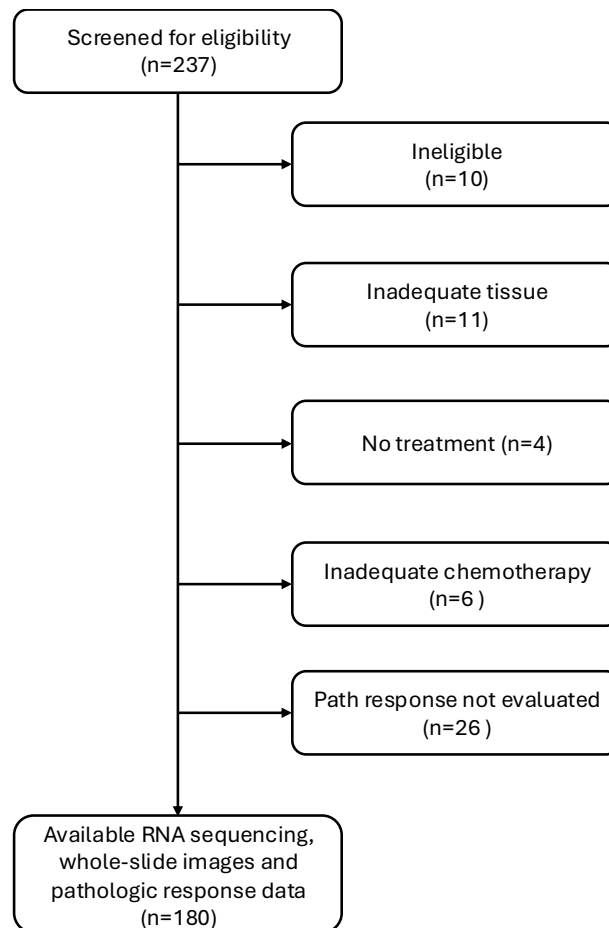
Supplementary Data 5. SHAP-based modality-level importance attribution for individual patients.

SHAP-based importance attribution values of the three modalities were computed using the modality-level proxy model for each individual patient in the hold-out test dataset.

Supplementary Filename: Supplementary_Data_5.xlsx

Index_Map: sheet (attribution plot) index/number to table (attribution plot) name mapping.

CONSORT Diagram



Our study conducted a secondary analysis of the trial data. Please refer to the original publication¹ for the primary trial data and CONSORT checklist.

1. A Randomized Phase II Study of Coexpression Extrapolation (COXEN) with Neoadjuvant Chemotherapy for Bladder Cancer. in *SWOG S1314; NCT02177695*).