

Transcriptome-based Deep Learning Model for Predicting Gemcitabine and Cisplatin Chemotherapy Response in Urothelial Carcinoma: Development and External Validation

JUWON KANG^{1,2*}, HYUN JUNG LEE^{3*}, SANG-BO OH⁴, JONG KIL NAM⁵, TAE UN KIM⁶, HWASEONG RYU⁶, YONG KAN KI⁷, JIHOON KANG¹, YI RANG KIM¹, JEONG HOON LEE⁸, JUNJEONG CHOI², YUN JEONG HONG⁹ and KWONOH PARK^{3,10}

¹ONCOCROSS Co., Ltd., Seoul, Republic of Korea;

²Yonsei Institute of Pharmaceutical Sciences, College of Pharmacy, Yonsei University, Incheon, Republic of Korea;

³Department of Pathology, Pusan National University Yangsan Hospital, Pusan National University School of Medicine, Yangsan, Republic of Korea;

⁴Medical Oncology and Hematology, Department of Internal medicine, Pusan National University Yangsan Hospital, Pusan National University School of Medicine, Yangsan, Republic of Korea;

⁵Department of Urology, Pusan National University Yangsan Hospital, Pusan National University School of Medicine, Yangsan, Republic of Korea;

⁶Department of Radiology, Pusan National University Yangsan Hospital, Pusan National University School of Medicine, Yangsan, Republic of Korea;

⁷Department of Radiation Oncology, Pusan National University Yangsan Hospital, Pusan National University School of Medicine, Yangsan, Republic of Korea;

⁸Department of Radiology, Stanford University School of Medicine, Stanford, CA, U.S.A.;

⁹Department of Neurology, Uijeongbu St. Mary's Hospital, Catholic University of Korea, Seoul, Republic of Korea;

¹⁰Division of Hematology & Medical Oncology, Department of Internal Medicine, Dongguk University Ilsan Hospital, Goyang, Republic of Korea

Abstract

Background/Aim: Chemotherapy with gemcitabine and cisplatin remains the cornerstone of treatment for advanced urothelial carcinoma (UC), yet response rates vary significantly among patients. Predicting treatment response is crucial to avoid unnecessary toxicity and optimize therapeutic strategies. This study aims to develop a deep learning model leveraging RNA sequencing data to predict chemotherapy response in UC patients.

Materials and Methods: We developed a deep learning model using RNA sequencing gene expression data from The Cancer Genome Atlas (TCGA) and Gene Expression Omnibus to predict chemotherapy (gemcitabine and cisplatin)

continued

*These Authors contributed equally to this work.

 Kwonoh Park, Division of Hematology & Medical Oncology, Department of Internal Medicine, Dongguk University Ilsan Hospital, 27 Dongguk-ro, Ilsandong-gu, Goyang-si, Gyeonggi-do 10326, Republic of Korea. Tel: +82 0319617096, e-mail: parkkoh@daum.net

Received January 14, 2026 | Revised February 15, 2026 | Accepted February 24, 2026



This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.
©2026 The Author(s). Anticancer Research is published by the International Institute of Anticancer Research.

response in UC patients. The model was externally validated using an independent cohort from the Pusan National University Yangsan Hospital. Model interpretation was performed through gene ontology and survival analyses using predictions from TCGA samples not included in the training set.

Results: The deep learning model demonstrated excellent predictive performance, achieving 94.7% accuracy in the training dataset and 90.0% accuracy in external validation. Gene ontology analysis revealed four key functional clusters associated with chemotherapy response: DNA damage response, cell cycle regulation, kinesins/microtubule dynamics, and mitotic cytokinesis. Notably, the model showed significant prognostic value in early-stage, with predicted responders displaying markedly better survival outcomes ($p=0.019$).

Conclusion: Our transcriptome-based deep learning approach offers a promising computational strategy for predicting chemotherapy response in urothelial carcinoma. By integrating high-dimensional RNA-seq data and advanced machine learning techniques, we provide a potential decision-support tool for personalized treatment planning.

Keywords: Urothelial carcinoma, cisplatin, gemcitabine, chemotherapy response, deep learning, RNA sequencing, precision oncology.

Introduction

Urothelial carcinoma (UC), also known as bladder cancer, is one of the most common urological cancers. Over 90% of UC cases originate in the uroepithelium (1). New treatment agents for UC, including immune checkpoint inhibitors, fibroblast growth factor receptor inhibitors, and enfortumab vedotin have demonstrated efficacy and are being applied in clinical practice (2-4). Nevertheless, cisplatin-based chemotherapy (CBCT), particularly with gemcitabine and cisplatin (GC), remains the most important first-line chemotherapy option for advanced UC (5, 6) and is used in neoadjuvant chemotherapy (NAC) for localized UC (7). While CBCT has a response rate of nearly 60%, it can also cause toxicity without efficacy in some patients (5). Predicting treatment response to chemotherapy is crucial for developing treatment strategies that minimize unnecessary chemotherapy, particularly in the NAC setting. Avoiding unnecessary chemotherapy helps preserve quality of life and prevents delays in radical cystectomy.

The heterogeneity of UC prognoses is due to the varying aggressiveness and proliferative capacity of cancer cells, and this nature is known to be determined by genetic information in UC (8, 9). However, identification of patient candidates at highest risk for recurrence or progression is currently based

on clinical information that may not reflect the entire biology of UC (10). The specific effects of genetic alterations that have been identified in UC need to be explored further.

Several studies have been conducted to identify molecular biomarkers for prognostic or predictive purposes in UC (11-13). Prior research has primarily focused on DNA repair genes as predictive biomarkers for CBCT response (14-17). However, these studies were limited to specific target molecules or pathways (18-20), and despite advances in molecular profiling (11, 21), comprehensive studies integrating whole transcriptome analysis for chemotherapy response prediction remain limited. RNA sequencing (RNA-seq) data enables comprehensive transcriptome profiling, allowing simultaneous evaluation of expression levels across human genes, thereby providing an extensive dataset for integrated transcriptomic analysis (22, 23). Deep learning models excel at capturing complex relationships among multiple features within large-scale datasets. This capability makes them particularly effective for analyzing gene expression data and improving prediction accuracy for personalized treatment strategies (24-27).

In this study, we developed a deep learning model to predict GC response in bladder cancer patients and evaluated its performance through external validation using independently collected patient data (Figure 1).

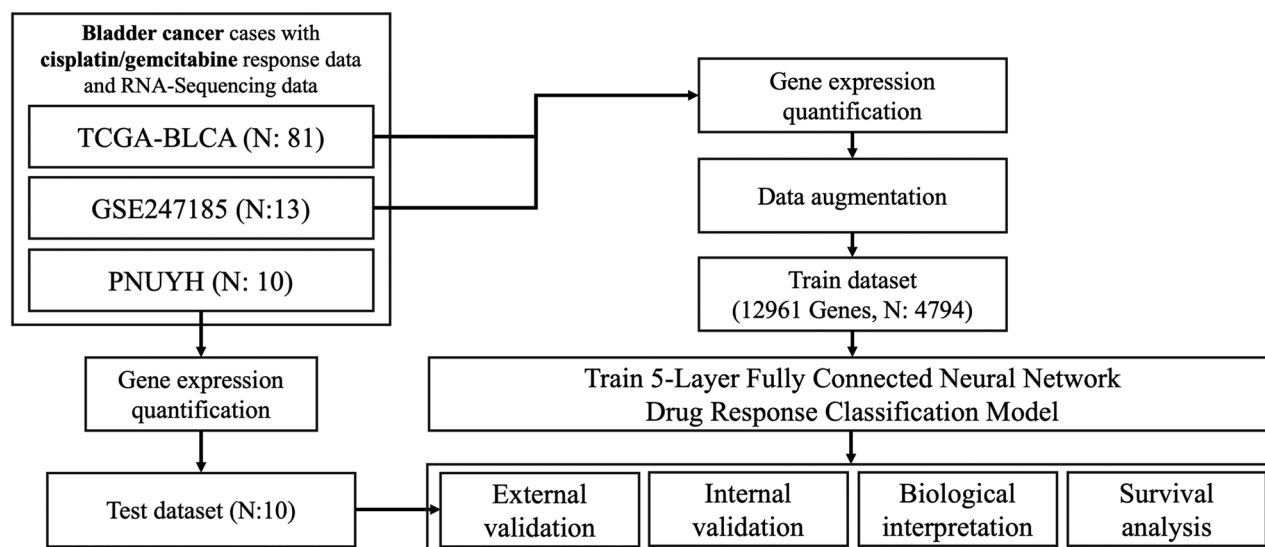


Figure 1. Overview of the study design and analytical workflow. Schematic representation of the study workflow. Bladder cancer cases with gemcitabine and cisplatin response data were collected from three sources: TCGA-BLCA (n=81), GSE247185 (n=13), and PNUYH (n=10). Gene expression data from TCGA and GSE datasets underwent quantification and augmentation to create the training dataset (12,961 genes, n=4,794), which was used to develop a 5-layer fully connected neural network for drug response classification. The PNUYH cohort served as an independent external validation set. The model was evaluated through internal validation, external validation, biological interpretation of predictive features, and survival analysis. PNUYH: Pusan National University Yangsan Hospital; TCGA-BLCA: the Cancer Genome Atlas Bladder Cancer cohort.

Materials and Methods

Data sources. We conducted a study using three independent datasets: The Cancer Genome Atlas (TCGA) bladder cancer cohort, Gene Expression Omnibus (GEO) dataset (GSE247185), and a clinical cohort from Pusan National University Yangsan Hospital (PNUYH). The study was approved by the Institutional Review Board of PNUYH (IRB Number: 05-2020-074).

Training dataset preparation. Primary training data were obtained from TCGA bladder cancer cohort using TCGA biolinks (version 2.30.4) in R (version 4.3.3). We extracted transcriptome quantification data (Transcripts Per Million, TPM) and matched clinical data from patients with bladder cancer who received GC regimen. Treatment response was categorized as responders (complete response or partial response) or non-responders (stable disease or progressive disease). Additional training data were acquired from GSE247185 (GSE), which included RNA sequencing data

from 13 pre-treatment formalin-fixed paraffin-embedded (FFPE) tumor biopsies of muscle-invasive bladder cancer patients who underwent neoadjuvant cisplatin-based chemotherapy. Response in this dataset was determined by pathologic complete response (pCR) status after treatment. Gene expression values were normalized using log2 transformation of TPM values. To enhance model robustness, we implemented data augmentation by generating 50-fold more samples than the original dataset, where each new sample was created by introducing a 5% random variation to individual gene expression values while preserving the underlying biological signal (28).

Internal validation. To assess model performance and stability, we implemented a 5-fold stratified cross-validation strategy. The augmented training dataset was divided into five equally sized folds using StratifiedKFold (scikit-learn, version 1.2.2), maintaining the same proportion of responders and non-responders in each fold. The stratification ensured balanced representation of

response classes across all folds. For each iteration, four folds of dataset were used for training and one-fold for validation, with random shuffling. Model performance was evaluated using multiple metrics including accuracy, sensitivity, specificity, precision, and F1-score. The receiver operating characteristic (ROC) curves and area under the curve (AUC) values were calculated to assess discriminative ability. Additionally, we monitored the training and validation loss curves across epochs to ensure proper model convergence and absence of overfitting.

External validation dataset preparation. The external validation cohort included patients diagnosed with advanced UC at PNUYH (from 2016 to 2019). Inclusion criteria were receipt of palliative GC chemotherapy, available response assessment based on RECIST version 1.1 criteria, and sufficient FFPE tissue for RNA-seq. Total RNA was extracted from FFPE tumor samples using the RNeasy FFPE kit (Qiagen, Hilden, Germany) and quantified using NanoDrop (ThermoFisher Scientific, Waltham, MA, USA). RNA-seq was performed on the Illumina platform using the SureSelectXT RNA Direct Reagent Kit (Agilent, Santa Clara, CA, USA) for paired-end sequencing. Raw sequencing data underwent quality control and trimming using Trimmomatic-0.39.1 (29). Sequence alignment was performed using STAR 2.7.8a (30), followed by expression quantification to TPM using RSEM 1.3.3 (31). Gene expression values were normalized using log2 transformation of TPM values. Patients were classified as responders (complete or partial response), or non-responders (stable or progressive disease) based on their treatment response.

Deep learning model architecture and training. Using gene expression data, we constructed a fully connected neural network (FCNN), a fundamental deep learning architecture (32, 33), to predict patient response to chemotherapy. The network architecture consisted of an input layer with 12,961 nodes corresponding to gene features, followed by five fully connected hidden layers with decreasing dimensions (1,024, 512, 128, 32 nodes), and a single output node with sigmoid activation. This

architecture resulted in approximately 13.9 million trainable parameters. To mitigate potential batch effects and internal covariate shift arising from diverse data sources (TCGA, GSE, and PNU) and different sample preservation methods (Fresh Frozen vs. FFPE), hidden layers employed layer normalization and SiLU activation functions (34, 35). Layer normalization was specifically chosen over batch normalization to ensure stable feature representation across heterogeneous datasets by normalizing each sample independently. In addition, a dropout layer was implemented before the first dense layer to prevent overfitting (36). The model was implemented in PyTorch (2.5.1+cu124) and trained using stochastic gradient descent optimization with a dynamic learning rate scheduler. Binary cross-entropy was used as the loss function, and training was conducted using mini-batch gradient descent. To ensure robust performance and assess model stability, the training procedure was done by ensemble model which was repeated 100 times with independent initializations (37).

Biological interpretation of the prediction model. To interpret the complex relationships between gene expressions and chemotherapy response in the prediction model, we performed correlation analysis between model predictions and gene expression levels using the TCGA bladder cancer cohort. The absolute values of Pearson correlation coefficients were calculated between predicted probabilities and expression levels of individual genes. The top 300 genes showing the highest absolute correlation values were selected for functional enrichment analysis. Gene ontology analysis was performed using DAVID Bioinformatics Resources 6.8, focusing on two annotation categories: Gene Ontology Biological Process (GOTERM_BP_DIRECT) and UniProt Keywords Biological Process (UP_KW_BIOLOGICAL_PROCESS) (38). Representative genes were selected from each enriched cluster based on their presence in multiple significant ontology terms. The relationships between genes and enriched ontology terms were visualized using GOplot package in R (39).

Prognostic value of model predictions. We conducted survival analysis to evaluate the prognostic value of our model's predictions. Using the TCGA bladder cancer cohort samples not included in the training set, patients were stratified into two groups based on their predicted response probabilities. Kaplan-Meier survival analysis was performed using the survminer package in R to compare overall survival between the predicted responder and non-responder groups. Statistical significance was assessed using the log-rank test, and survival curves were generated with risk tables showing the number of patients at risk at different time points.

Results

Dataset characteristics. Our study utilized two distinct datasets for training: the TCGA bladder cancer cohort and the GSE dataset. The training set included a total of 94 samples, with 81 samples from TCGA (37 non-responders, 44 responders) and 13 samples from GSE (5 non-responders, 8 responders). For external validation, we sampled an independent cohort of 10 stage IV bladder cancer patients from PNUYH, consisting of 7 responders and 3 non-responders to GC chemotherapy. The baseline characteristics of PNUYH patients are shown in Table I. To enhance model robustness, we implemented data augmentation, resulting in a total of 4,794 augmented samples used for model development. For input feature, we identified 12,961 genes that were consistently detected across all three datasets, ensuring comprehensive transcriptomic representation while maintaining cross-dataset compatibility.

Model performance in internal and external validation. The model training converged successfully over 200 epochs, with consistent decreases in loss values for both training and validation sets (Figure 2B, E). In internal validation, the model demonstrated excellent discriminative ability with ROC curves showing high performance across multiple cross-validation folds (Figure 2A). We analyzed precision, recall, and F1-score metrics across different threshold values and determined 0.4 as the optimal prediction threshold (Figure 2C). The 5-fold cross-validation results showed

Table I. *Clinicopathologic features of patients with advanced urothelial carcinoma (N=10).*

Characteristics	Number (%)
Age (years)	
Median (Range)	65.5 (52-76)
Sex	
Male	10 (100%)
Female	0
ECOG Performance status	
0-1	8 (80%)
2	2 (20%)
Primary tumor site	
Bladder	6 (60%)
UTUC	4 (40%)
Best response	
Partial response	7 (70%)
Stable disease	0
Progressive disease	3 (30%)

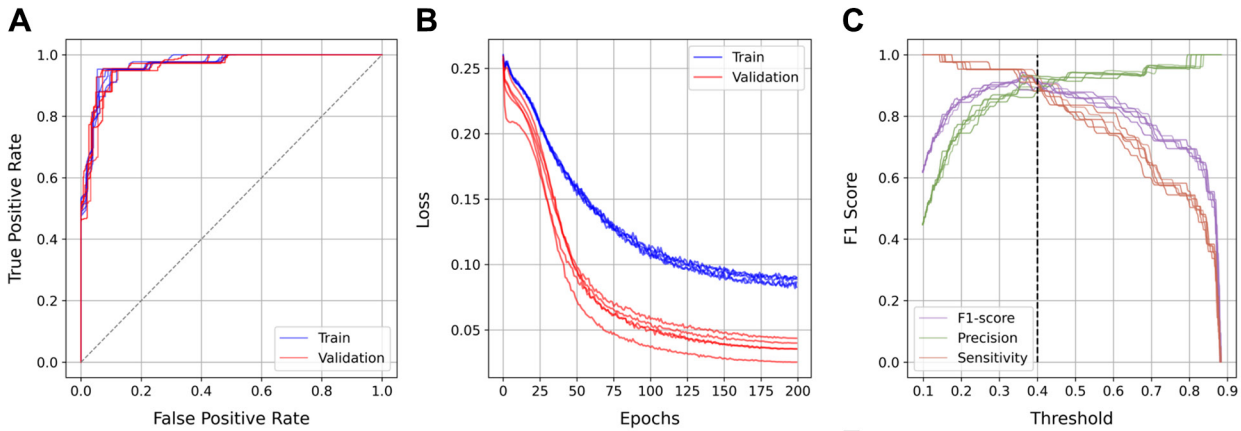
ECOG, Eastern Cooperative Oncology Group; UTUC, upper tract urothelial carcinoma.

consistent performance with average accuracy, sensitivity, specificity, precision, and F1-scores of 0.914, 0.906, 0.920, 0.902, and 0.904 for training, and 0.905, 0.892, 0.915, 0.895, and 0.893 for validation, respectively (Figure 2F).

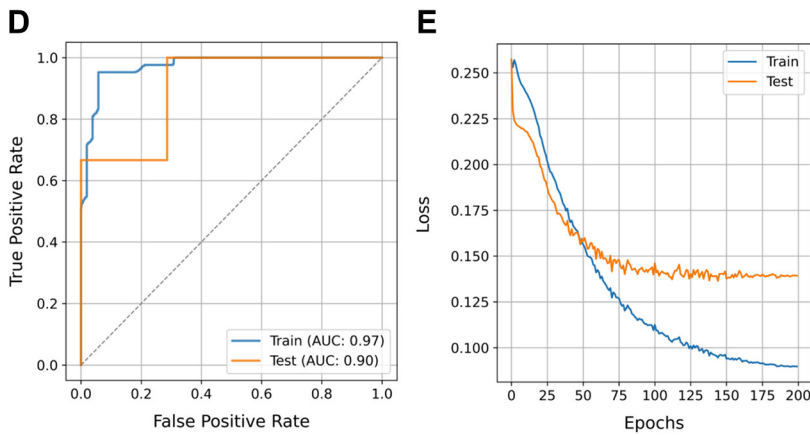
In external validation using an independent cohort of stage IV bladder cancer patients (PNUYH, $n=10$), the model maintained robust performance with an area under the receiver operating characteristic curve (AUROC) of 0.90 (Figure 2D). When applying the optimal threshold of 0.4, the model achieved an accuracy of 90.0% (95% CI=55.5-99.7), with a sensitivity for non-responders of 66.7% (95% CI=9.4-99.2), specificity of 100.0% (95% CI=59.0-100.0), and precision of 1.000 (Figure 2G), demonstrating strong generalizability in identifying treatment response in a completely independent patient population.

Biological functions associated with chemotherapy response prediction. Gene enrichment analysis of the top 300 correlated genes revealed four major functional clusters. The significance of enriched terms was visualized using negative log2-transformed p -values (Figure 3A). The first cluster contained terms related to DNA damage response and repair, with DNA damage and DNA repair being significantly enriched across multiple ontology databases.

[Internal Validation]



[External Validation]



F

Fold	Dataset	Accuracy	Sensitivity	Specificity	Precision	F1-score
1	Train	0.919	0.930	0.910	0.893	0.911
	Validation	0.898	0.923	0.877	0.859	0.890
2	Train	0.894	0.881	0.905	0.882	0.882
	Validation	0.891	0.879	0.900	0.877	0.878
3	Train	0.923	0.901	0.941	0.925	0.913
	Validation	0.923	0.890	0.949	0.934	0.911
4	Train	0.905	0.911	0.901	0.881	0.896
	Validation	0.902	0.881	0.919	0.898	0.889
5	Train	0.928	0.906	0.946	0.931	0.918
	Validation	0.910	0.888	0.928	0.909	0.898
Average	Train	0.914	0.906	0.920	0.902	0.904
	Validation	0.905	0.892	0.915	0.895	0.893

G

Dataset	Accuracy	Sensitivity	Specificity	Precision	F1-score
Train	0.947	0.952	0.942	0.930	0.941
Test	0.900	0.667	1.000	1.000	0.800

Test Dataset(N:10)	Predicted Responder	Predicted Non-responder
Actual Responder	7	0
Actual Non-responder	1	2

Figure 2. Performance evaluation of the deep learning model in internal and external validation. (A–C) Internal validation results: (A) ROC curves are presented for the training and internal validation sets; (B) Training and validation loss curves over 200 epochs showing stable convergence; (C) Precision, recall, and F1-score metrics across different threshold values with optimal threshold (0.4) indicated by the vertical dashed line. (D–G) External validation results: (D) ROC curves for training and external test sets with respective AUROCs; (E) Loss curves during external validation showing good generalization; (F) Detailed performance metrics for each fold during 5-fold cross-validation; (G) Final performance metrics and confusion matrix of the model in training and external validation datasets. AUROC: Area under the receiver operating characteristic curve; ROC, receiver operating characteristic.

Key genes in this cluster included *BARD1*, *BRCA2*, *RIF1*, *SMC3*, *USP1* and *MSH6*. The second cluster comprised cell cycle regulation terms including cell cycle, cell division, and mitosis. This cluster included genes *KIF11*, *CDK8*, *PRC1* and *CENPE*. The third cluster was associated with kinesins and microtubule-based movement, featuring genes *KIF11*, *CENPE*, *KIF4B*, *KIF1B* and *KIF23*. The fourth cluster contained terms related to mitotic cytokinesis and RHOA GTPase cycle regulation. The relationships between

representative genes and their associated biological processes were visualized using a chord diagram (Figure 3B), illustrating how many genes were involved in multiple functional pathways, particularly connecting cell cycle regulation with DNA repair mechanisms.

Prognostic significance and stage-specific predictions. We next investigated whether our model's predictions correlated with patient survival outcomes across different disease

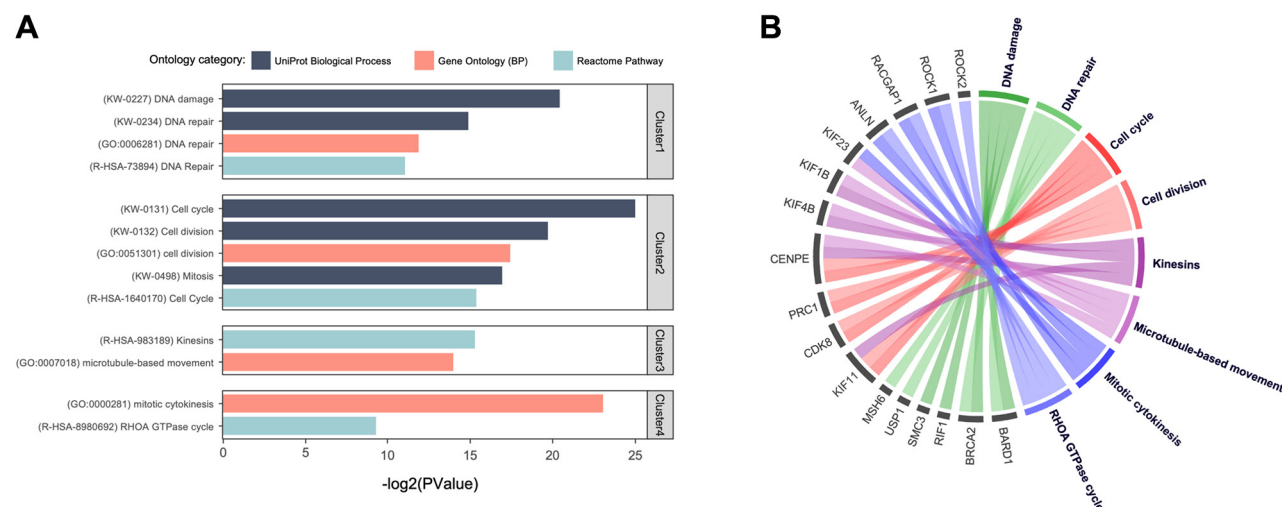


Figure 3. Gene ontology enrichment analysis of chemotherapy response-associated genes. (A) Bar plots showing enriched biological process terms in four major functional clusters. The x-axis represents negative log₂-transformed p-Values, and bars are colored by gene ontology category. (B) Chord diagram illustrating the relationships between representative genes and their associated biological processes.

stages (Figure 4). Kaplan-Meier analysis revealed that in early-stage (Stage I/II) bladder cancer patients, the model's predictions demonstrated significant prognostic value ($p=0.019$), with predicted responders showing markedly better survival than predicted non-responders (Figure 4A). This prognostic distinction was not statistically significant in Stage III ($p=0.69$, Figure 4B) or Stage IV patients ($p=0.2$, Figure 4C). Notably, we observed a gradual increase in prediction values across disease stages, with the highest prediction scores observed in Stage IV patients (Figure 4D). The median prediction values were 0.275, 0.314, and 0.527 for Stage I/II, Stage III, and Stage IV patients, respectively, while mean prediction values increased from 0.376 in Stage I/II to 0.414 in Stage III and 0.514 in Stage IV. This stage-dependent pattern suggests that our model captures molecular features that may be more prevalent in advanced disease, potentially reflecting differences in tumor biology and treatment response mechanisms across disease stages.

Discussion

Our deep learning model demonstrated strong predictive performance, achieving 94.7% accuracy in the training

dataset and 90.0% accuracy in external clinical validation from PNUYH. Notably, the model maintained 100.0% specificity, indicating its reliability in identifying true responders. The high accuracy in an independent cohort, derived from FFPE tissue samples, underscores its robustness across different patient populations and clinical settings. To justify the use of a deep learning architecture (FCNN) despite its complexity, we conducted a benchmarking analysis against traditional machine learning models, including Logistic Regression, Random Forest, and Support vector machines. Our FCNN significantly outperformed these simpler models, which showed limited generalizability on the external cohort (Table II). This suggests that the multi-layered deep learning approach is essential for capturing the high-order, non-linear interactions within the high-dimensional transcriptomic space (12,961 genes) that simpler models fail to distill. Our comprehensive transcriptomic approach offers advantages by integrating information from thousands of genes, capturing complex gene interaction patterns that contribute to treatment response, rather than relying on single-gene markers that may not fully reflect the heterogeneous mechanisms of gemcitabine and cisplatin resistance.

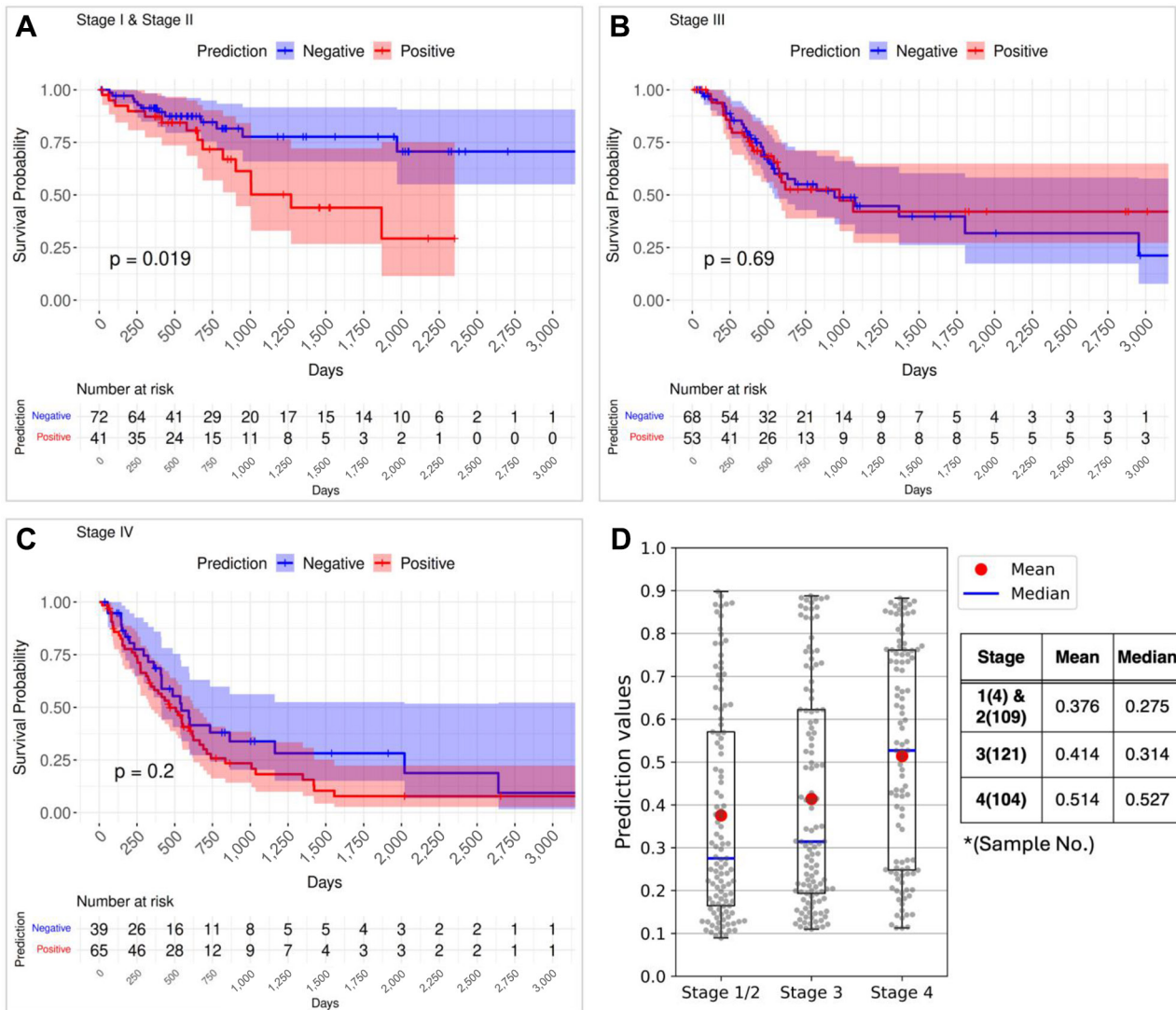


Figure 4. Stage-specific survival analysis and prediction distributions. (A-C) Kaplan-Meier survival curves stratified by model predictions (Negative: predicted responders, Positive: predicted non-responders) across different disease stages: (A) Stage I/II patients showing significant survival difference ($p=0.019$); (B) Stage III patients showing no significant difference in survival outcomes ($p=0.69$); (C) Stage IV patients showing a non-significant trend toward better survival in predicted responders ($p=0.2$). Numbers of patients at risk are shown below each graph. (D) Box plot showing the distribution of prediction values across disease stages, with individual data points, medians (blue lines), and means (red dots) displayed. Higher prediction values indicate greater likelihood of non-response to chemotherapy.

Functional analysis of the top 300 associated genes revealed four major clusters: DNA damage response, cell cycle regulation, kinesins/microtubule dynamics, and mitotic cytokinesis, all of which are closely linked to the mechanism of action of chemotherapy. Among these, genes involved in DNA damage response, including

BARD1, *BRCA2*, *RIF1*, and *USP1*, emerged as key predictive markers, reinforcing the established role of DNA repair pathways in chemotherapy sensitivity (40-43). These findings reinforce the role of DNA repair pathways in chemotherapy sensitivity, suggesting that transcriptomic profiling of these genes could enhance response prediction.

Particularly intriguing was our observation of increasing prediction scores with advancing disease stage, suggesting that molecular features associated with chemotherapy resistance become more prevalent in advanced tumors. This progressive molecular shift may explain why early-stage tumors showed significant survival stratification based on our model's predictions, while advanced tumors exhibited generally higher resistance signatures regardless of individual response outcomes.

The ability to predict chemotherapy response before treatment initiation has significant clinical implications, particularly in the neoadjuvant setting where unnecessary chemotherapy can delay definitive surgery and impact quality of life (7, 44). In the field of UC, various deep learning models have been developed for early detection (45), and RNA-based approaches have also been explored for therapeutic target identification (46). Our model's high specificity in identifying non-responders suggests it could be valuable in helping clinicians avoid ineffective treatments in patients unlikely to benefit from chemotherapy. The prognostic value demonstrated in early-stage patients further indicates potential utility in risk stratification and treatment planning. For patients predicted to have high resistance to GC regimen, alternative approaches such as immune checkpoint inhibitors or antibody-drug conjugates could be considered earlier in the treatment course (47, 48). Additionally, the model's performance on FFPE samples—the standard specimen type in clinical practice—enhances its translational potential, suggesting feasibility for integration into routine pathological workflows after further validation. The consistent performance across different gene expression platforms also indicates robust technical transferability, which is essential for clinical implementation.

Despite the promising results, our study has several limitations. First, while our model demonstrated high accuracy, the sample size remains modest ($n=94$ for discovery, $n=10$ for PNUYH validation). To ensure stable model convergence given the high-dimensional transcriptomic features and limited sample size, data augmentation was applied to the discovery cohort prior to

Table II. Performance comparison between the proposed deep learning model and traditional machine learning algorithms in the external validation cohort ($N=10$).

Model	Accuracy	AUROC
Deep learning (Proposed)	0.9	0.9
Logistic regression	0.7	0.619
Random forest	0.6	0.714
Support vector machine	0.7	0.429

AUROC: Area under the receiver operating characteristic curve.

cross-validation. While we acknowledge that this strategy may lead to potential data leakage and subsequent inflation of internal validation metrics, the robust performance observed in the completely independent, non-augmented PNUYH external validation cohort (accuracy 90.0%, AUROC 0.90) confirms the model's genuine generalizability and clinical utility. This indicates that the model successfully learned robust biological features associated with treatment response rather than overfitting to augmented noise. Nonetheless, prospective multicenter studies with larger cohorts are warranted to further validate these findings. Second, our current model relies on bulk RNA sequencing, which may not fully capture tumor heterogeneity or the influence of the tumor microenvironment on treatment response. Integration of spatial transcriptomics or single-cell approaches could provide deeper insights into resistance mechanisms. Third, the model's interpretability, while improved through our functional analysis, still presents challenges for direct clinical translation. Future work should focus on developing more transparent predictive models that maintain performance while providing clearer biological insights.

Conclusion

This study demonstrates the potential of deep learning-based transcriptomic analysis in predicting GC chemotherapy response in UC patients. By leveraging high-dimensional RNA-seq data, our model effectively identified molecular features associated with treatment outcomes and maintained high predictive accuracy in external validation.

These findings highlight the feasibility of integrating transcriptomic deep learning models into clinical workflows, emphasizing their potential as a clinically viable decision-support tool to guide personalized treatment strategies.

Data Availability Statement

The sequencing datasets have been deposited in the Gene Expression Omnibus (GEO) repository under accession number GSE290327.

Conflicts of Interest

Juwon Kang, Jihoon Kang, and Yi Rang Kim are employees of ONCOCROSS Co., Ltd.

Authors' Contributions

Conceptualization, K.P., YR.K.; methodology, J.K. and HJ.L.; software, JH.L.; validation, JH.L. and J.C.; formal analysis, J.K. and JH.L.; investigation, J.K. and J.C.; resources, SB.O., JK.N., TU.K., H.R., and YK.K.; data curation, J.K. HJ.L., J.K.; writing – original draft preparation, J.K.; writing – review and editing, HJ.L., SB.O., JK.N., TU.K., H.R., YK.K., YR.K., J.C., and K.P.; visualization, J.K. and JH.L.; supervision, J.C. and K.P.; funding acquisition, K.P. All Authors have read and agreed to the published version of the manuscript.

Funding

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) (No. 202107620001) and by a grant (21153MFDS601) from the Ministry of Food and Drug Safety in 2023.

Artificial Intelligence (AI) Disclosure

During the preparation of this manuscript, a large language model (ChatGPT, OpenAI) was used solely for language editing and stylistic refinement in select paragraphs. No content related to the generation, analysis,

or interpretation of research data were produced by generative AI. All scientific content and conclusions were developed and verified by the authors. No figures or visual data were generated or modified using generative AI or machine learning-based image enhancement tools.

References

- 1 Wasco MJ, Daignault S, Zhang Y, Kunju LP, Kinnaman M, Braun T, Lee CT, Shah RB: Urothelial carcinoma with divergent histologic differentiation (mixed histologic features) predicts the presence of locally advanced bladder cancer when detected at transurethral resection. *Urology* 70(1): 69-74, 2007. DOI: 10.1016/j.urology.2007.03.033
- 2 Loriot Y, Necchi A, Park SH, Garcia-Donas J, Huddart R, Burgess E, Fleming M, Rezazadeh A, Mellado B, Varlamov S, Joshi M, Duran I, Tagawa ST, Zakharia Y, Zhong B, Stuyckens K, Santiago-Walker A, De Porre P, O'Hagan A, Avadhani A, Siefker-Radtke AO, BLC2001 Study Group: Erdafitinib in locally advanced or metastatic urothelial carcinoma. *N Engl J Med* 381(4): 338-348, 2019. DOI: 10.1056/NEJMoa1817323
- 3 Powles T, Park SH, Voog E, Caserta C, Valderrama BP, Gurney H, Kalofonos H, Radulović S, Demey W, Ullén A, Loriot Y, Sridhar SS, Tsuchiya N, Kopyltsov E, Sternberg CN, Bellmunt J, Aragon-Ching JB, Petrylak DP, Laliberte R, Wang J, Huang B, Davis C, Fowst C, Costa N, Blake-Haskins JA, di Pietro A, Grivas P: Avelumab maintenance therapy for advanced or metastatic urothelial carcinoma. *N Engl J Med* 383(13): 1218-1230, 2020. DOI: 10.1056/NEJMoa2002788
- 4 Powles T, Rosenberg JE, Sonpavde GP, Loriot Y, Durán I, Lee JL, Matsubara N, Vulsteke C, Castellano D, Wu C, Campbell M, Matsangou M, Petrylak DP: Enfortumab vedotin in previously treated advanced urothelial carcinoma. *N Engl J Med* 384(12): 1125-1135, 2021. DOI: 10.1056/NEJMoa2035807
- 5 von der Maase H, Hansen SW, Roberts JT, Dogliotti L, Oliver T, Moore MJ, Bodrogi I, Albers P, Knuth A, Lippert CM, Kerbrat P, Sanchez Rovira P, Wersall P, Cleall SP, Roychowdhury DF, Tomlin I, Visseren-Grul CM, Conte PF: Gemcitabine and cisplatin versus methotrexate, vinblastine, doxorubicin, and cisplatin in advanced or metastatic bladder cancer: results of a large, randomized, multinational, multicenter, phase III study. *J Clin Oncol* 18(17): 3068-3077, 2000. DOI: 10.1200/JCO.2000.18.17.3068
- 6 Minato A, Murooka K, Okumura Y, Takaba T, Higashijima K, Nagata Y, Tomisaki I, Harada K, Fujimoto N: Efficacy of platinum-based chemotherapy in patients with metastatic urothelial carcinoma with variant histology. *In Vivo* 38(2): 873-880, 2024. DOI: 10.21873/in vivo.13513
- 7 Pfister C, Gravis G, Fléchon A, Soulié M, Guy L, Laguerre B, Mottet N, Joly F, Allory Y, Harter V, Culine S, VESPER Trial Investigators: Randomized phase III trial of dose-dense

- methotrexate, vinblastine, doxorubicin, and cisplatin, or gemcitabine and cisplatin as perioperative chemotherapy for patients with muscle-invasive bladder cancer. Analysis of the GETUG/AFU V05 VESPER trial secondary endpoints: chemotherapy toxicity and pathological responses. *Eur Urol* 79(2): 214-221, 2021. DOI: 10.1016/j.eururo.2020.08.024
- 8 Cancer Genome Atlas Research Network: Comprehensive molecular characterization of urothelial bladder carcinoma. *Nature* 507(7492): 315-322, 2014. DOI: 10.1038/nature12965
- 9 Kim J, Akbani R, Creighton CJ, Lerner SP, Weinstein JN, Getz G, Kwiatkowski DJ: Invasive bladder cancer: genomic insights and therapeutic promise. *Clin Cancer Res* 21(20): 4514-4524, 2015. DOI: 10.1158/1078-0432.CCR-14-1215
- 10 International Bladder Cancer Nomogram Consortium, Bochner BH, Kattan MW, Vora KC: Postoperative nomogram predicting risk of recurrence after radical cystectomy for bladder cancer. *J Clin Oncol* 24(24): 3967-3972, 2006. DOI: 10.1200/JCO.2005.05.3884
- 11 Choi W, Ochoa A, McConkey DJ, Aine M, Höglund M, Kim WY, Real FX, Kiltie AE, Milsom I, Dyrskjöt L, Lerner SP: Genetic alterations in the molecular subtypes of bladder cancer: Illustration in the Cancer Genome Atlas Dataset. *Eur Urol* 72(3): 354-365, 2017. DOI: 10.1016/j.eururo.2017.03.010
- 12 Buttiglieri C, Tucci M, Vignani F, Scagliotti GV, Di Maio M: Molecular biomarkers to predict response to neoadjuvant chemotherapy for bladder cancer. *Cancer Treat Rev* 54: 1-9, 2017. DOI: 10.1016/j.ctrv.2017.01.002
- 13 Sjö Dahl G, Abrahamsson J, Holmsten K, Bernardo C, Chebil G, Eriksson P, Johansson I, Kollberg P, Lindh C, Lövgren K, Marzouka N, Olsson H, Höglund M, Ullén A, Liedberg F: Different responses to neoadjuvant chemotherapy in urothelial carcinoma molecular subtypes. *Eur Urol* 81(5): 523-532, 2022. DOI: 10.1016/j.eururo.2021.10.035
- 14 Plimack ER, Dunbrack RL, Brennan TA, Andrade MD, Zhou Y, Serebriiskii IG, Slifker M, Alpaugh K, Dulaimi E, Palma N, Hoffman-Censits J, Bilusic M, Wong YN, Kutikov A, Viterbo R, Greenberg RE, Chen DY, Lallas CD, Trabulsi EJ, Yelensky R, McConkey DJ, Miller VA, Golem EA, Ross EA: Defects in DNA repair genes predict response to neoadjuvant cisplatin-based chemotherapy in muscle-invasive bladder cancer. *Eur Urol* 68(6): 959-967, 2015. DOI: 10.1016/j.eururo.2015.07.009
- 15 Van Allen EM, Mouw KW, Kim P, Iyer G, Wagle N, Al-Ahmadie H, Zhu C, Ostrovnaya I, Kryukov GV, O'Connor KW, Sfakianos J, Garcia-Grossman I, Kim J, Guancial EA, Bambury R, Bahl S, Gupta N, Farlow D, Qu A, Signoretti S, Barletta JA, Reuter V, Boehm J, Lawrence M, Getz G, Kantoff P, Bochner BH, Choueiri TK, Bajorin DF, Solit DB, Gabriel S, D'Andrea A, Garraway LA, Rosenberg JE: Somatic ERCC2 mutations correlate with cisplatin sensitivity in muscle-invasive urothelial carcinoma. *Cancer Discov* 4(10): 1140-1153, 2014. DOI: 10.1158/2159-8290.CD-14-0623
- 16 Liu D, Plimack ER, Hoffman-Censits J, Garraway LA, Bellmunt J, Van Allen E, Rosenberg JE: Clinical validation of chemotherapy response biomarker ERCC2 in muscle-invasive urothelial bladder carcinoma. *JAMA Oncol* 2(8): 1094-1096, 2016. DOI: 10.1001/jamaoncol.2016.1056
- 17 Iyer G, Balar AV, Milowsky MI, Bochner BH, Dalbagni G, Donat SM, Herr HW, Huang WC, Taneja SS, Woods M, Ostrovnaya I, Al-Ahmadie H, Arcila ME, Riches JC, Meier A, Bourque C, Shady M, Won H, Rose TL, Kim WY, Kania BE, Boyd ME, Cipolla CK, Regazzi AM, Delbeau D, McCoy AS, Vargas HA, Berger MF, Solit DB, Rosenberg JE, Bajorin DF: Multicenter prospective phase II trial of neoadjuvant dose-dense gemcitabine plus cisplatin in patients with muscle-invasive bladder cancer. *J Clin Oncol* 36(19): 1949-1956, 2018. DOI: 10.1200/JCO.2017.75.0158
- 18 Wang L, Wei J, Qian X, Yin H, Zhao Y, Yu L, Wang T, Liu B: ERCC1 and BRCA1 mRNA expression levels in metastatic malignant effusions is associated with chemosensitivity to cisplatin and/or docetaxel. *BMC Cancer* 8: 97, 2008. DOI: 10.1186/1471-2407-8-97
- 19 Bellmunt J, Paz-Ares L, Cuello M, Cecere FL, Albiol S, Guillem V, Gallardo E, Carles J, Mendez P, de la Cruz JJ, Taron M, Rosell R, Baselga J, Spanish Oncology Genitourinary Group: Gene expression of ERCC1 as a novel prognostic marker in advanced bladder cancer patients receiving cisplatin-based chemotherapy. *Ann Oncol* 18(3): 522-528, 2007. DOI: 10.1093/annonc/mdl435
- 20 Font A, Taron M, Gago JL, Costa C, Sánchez JJ, Carrato C, Mora M, Celiz P, Perez L, Rodríguez D, Gimenez-Capitan A, Quiroga V, Benlloch S, Ibarz L, Rosell R: BRCA1 mRNA expression and outcome to neoadjuvant cisplatin-based chemotherapy in bladder cancer. *Ann Oncol* 22(1): 139-144, 2011. DOI: 10.1093/annonc/mdq333
- 21 Knowles MA, Hurst CD: Molecular biology of bladder cancer: new insights into pathogenesis and clinical diversity. *Nat Rev Cancer* 15(1): 25-41, 2015. DOI: 10.1038/nrc3817
- 22 Wang Z, Gerstein M, Snyder M: RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* 10(1): 57-63, 2009. DOI: 10.1038/nrg2484
- 23 Conesa A, Madrigal P, Tarazona S, Gomez-Cabrero D, Cervera A, McPherson A, Szczesniak MW, Gaffney DJ, Elo LL, Zhang X, Mortazavi A: A survey of best practices for RNA-seq data analysis. *Genome Biol* 17: 13, 2016. DOI: 10.1186/s13059-016-0881-8
- 24 Min S, Lee B, Yoon S: Deep learning in bioinformatics. *Brief Bioinform* 18(5): 851-869, 2016. DOI: 10.1093/bib/bbw068
- 25 Angermueller C, Pärnamäa T, Parts L, Stegle O: Deep learning for computational biology. *Mol Syst Biol* 12(7): 878, 2016. DOI: 10.15252/msb.20156651
- 26 Zou J, Huss M, Abid A, Mohammadi P, Torkamani A, Telenti A: A primer on deep learning in genomics. *Nat Genet* 51(1): 12-18, 2019. DOI: 10.1038/s41588-018-0295-5
- 27 Ching T, Himmelstein DS, Beaulieu-Jones BK, Kalinin AA, Do BT, Way GP, Ferrero E, Agapow PM, Zietz M, Hoffman MM, Xie W, Rosen GL, Lengerich BJ, Israeli J, Lanchantin J, Woloszynek

- S, Carpenter AE, Shrikumar A, Xu J, Cofer EM, Lavender CA, Turaga SC, Alexandari AM, Lu Z, Harris DJ, DeCaprio D, Qi Y, Kundaje A, Peng Y, Wiley LK, Segler MHS, Boca SM, Swamidass SJ, Huang A, Gitter A, Greene CS: Opportunities and obstacles for deep learning in biology and medicine. *J R Soc Interface* 15(141): 20170387, 2018. DOI: 10.1098/rsif.2017.0387
- 28 Kircher M, Chludzinski E, Krepel J, Saremi B, Beineke A, Jung K: Augmentation of transcriptomic data for improved classification of patients with respiratory diseases of viral origin. *Int J Mol Sci* 23(5): 2481, 2022. DOI: 10.3390/ijms23052481
- 29 Bolger AM, Lohse M, Usadel B: Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30(15): 2114-2120, 2014. DOI: 10.1093/bioinformatics/btu170
- 30 Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR: STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29(1): 15-21, 2013. DOI: 10.1093/bioinformatics/bts635
- 31 Li B, Dewey CN: RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 12: 323, 2011. DOI: 10.1186/1471-2105-12-323
- 32 LeCun Y, Bengio Y, Hinton G: Deep learning. *Nature* 521(7553): 436-444, 2015. DOI: 10.1038/nature14539
- 33 Hornik K, Stinchcombe M, White H: Multilayer feedforward networks are universal approximators. *Neural Networks* 2(5): 359-366, 1989. DOI: 10.1016/0893-6080(89)90020-8
- 34 Ramachandran P, Zoph B, Le QV: Searching for activation functions. *arXiv: 171005941*, 2017.
- 35 Ba J, Kiros JR, Hinton GE: Layer normalization. *arXiv: 1607.06450*, 2016.
- 36 Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R: Dropout: A simple way to prevent neural networks from overfitting. *J Machine Learn Res* 15(56): 1929-1958, 2014.
- 37 Ganaie MA, Hu M, Malik AK, Tanveer M, Suganthan PN: Ensemble deep learning: A review. *Eng Appl Artif Intell* 115: 105151, 2022. DOI: 10.1016/j.engappai.2022.105151
- 38 Huang DW, Sherman BT, Lempicki RA: Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 4(1): 44-57, 2009. DOI: 10.1038/nprot.2008.211
- 39 Walter W, Sánchez-Cabo F, Ricote M: GOpilot: an R package for visually combining expression data with functional analysis. *Bioinformatics* 31(17): 2912-2914, 2015. DOI: 10.1093/bioinformatics/btv300
- 40 Albarakati N, Al-Ghamdi H, Al-Sowayan B, Alshareeda A: Homologous recombination mRNAs (RAD21, RAD50 and BARD1) have a potentially poor prognostic role in ERBB2-low bladder cancer patients. *Sci Rep* 13(1): 11738, 2023. DOI: 10.1038/s41598-023-38923-y
- 41 Taber A, Christensen E, Lamy P, Nordentoft I, Prip F, Lindskrog SV, Birkenkamp-Demtröder K, Okholm TLH, Knudsen M, Pedersen JS, Steiniche T, Agerbæk M, Jensen JB, Dyrskjød L: Molecular correlates of cisplatin-based chemotherapy response in muscle invasive bladder cancer by integrated multi-omics analysis. *Nat Commun* 11(1): 4858, 2020. DOI: 10.1038/s41467-020-18640-0
- 42 Mei Y, Peng C, Liu YB, Wang J, Zhou HH: Silencing RIF1 decreases cell growth, migration and increases cisplatin sensitivity of human cervical cancer cells. *Oncotarget* 8(63): 107044-107051, 2017. DOI: 10.18632/oncotarget.22315
- 43 Sonogo M, Pellarin I, Costa A, Vinciguerra GLR, Coan M, Kraut A, D'Andrea S, Dall'Acqua A, Castillo-Tong DC, Califano D, Losito S, Spizzo R, Couté Y, Vecchione A, Belletti B, Schiappacassi M, Baldassarre G: USP1 links platinum resistance to cancer cell dissemination by regulating Snail stability. *Sci Adv* 5(5): eaav3235, 2019. DOI: 10.1126/sciadv.aav3235
- 44 Yang M, Han J, Won H, Chae SW, Kim HS, Do SI: Automated prediction of pathological complete response to neoadjuvant chemotherapy in breast carcinoma using deep learning on pretreatment core needle biopsy samples. *Anticancer Res* 45(11): 5167-5176, 2025. DOI: 10.21873/anticancer.17856
- 45 Tung C, Lin S, Chang K, Xu Y, Chuang M, Chu Y: Light Bladder Net: Non-invasive bladder cancer prediction by weighted deep learning approaches and graphical data transformation. *Anticancer Res* 45(5): 1953-1964, 2025. DOI: 10.21873/anticancer.17572
- 46 Nahkuri S, Weidle UH: Bladder cancer: role of circular RNAs in oncogenesis, tumor suppression, and therapeutic target identification. *Cancer Genomics Proteomics* 22(5): 654-682, 2025. DOI: 10.21873/cgp.20528
- 47 Powles T, Catto JWF, Galsky MD, Al-Ahmadie H, Meeks JJ, Nishiyama H, Vu TQ, Antonuzzo L, Wiechno P, Atduiev V, Kann AG, Kim TH, Suárez C, Chang CH, Roghmann F, Özgüroğlu M, Eigel BJ, Oliveira N, Buchler T, Gadot M, Zakharia Y, Armstrong J, Gupta A, Hois S, van der Heijden MS, NIAGARA Investigators: Perioperative durvalumab with neoadjuvant chemotherapy in operable bladder cancer. *N Engl J Med* 391(19): 1773-1786, 2024. DOI: 10.1056/NEJMoa2408154
- 48 Galsky MD, Hoimes CJ, Necchi A, Shore N, Witjes JA, Steinberg G, Bedke J, Nishiyama H, Fang X, Kataria R, Sbar E, Jia X, Siefker-Radtke A: Perioperative pembrolizumab therapy in muscle-invasive bladder cancer: Phase III KEYNOTE-866 and KEYNOTE-905/EV-303. *Future Oncol* 17(24): 3137-3150, 2021. DOI: 10.2217/fon-2021-0273