

Transcriptome-based high-frequency recurrence index predicts frequent recurrence in non-muscle-invasive bladder cancer after Bacillus Calmette-Guérin therapy

Received: 03 Sep 2025

Accepted: 23 Feb 2026

Published online: 27 February 2026

Cite this article as: Baek, S., Byun, Y., Piao, X. *et al.* Transcriptome-based high-frequency recurrence index predicts frequent recurrence in non-muscle-invasive bladder cancer after Bacillus Calmette-Guérin therapy. *BMC Med* (2026). <https://doi.org/10.1186/s12916-026-04745-8>

Seung-Woo Baek, Young Byun, Xuan-Mei Piao, Jee-Woo Seo, Il-San Jeong, Otabek Kamarov, Seongmin Moon, Kyeong Kim, Ho Kang, Won Kim, Yung Choi, Sung-Kwon Moon, Wun-Jae Kim, Young Choi, Seon-Young Kim, Seon-Kyu Kim & Seok Yun

We are providing an unedited version of this manuscript to give early access to its findings. Before final publication, the manuscript will undergo further editing. Please note there may be errors present which affect the content, and all legal disclaimers apply.

If this paper is publishing under a Transparent Peer Review model then Peer Review reports will publish with the final article.

Transcriptome-based High-Frequency Recurrence Index Predicts Frequent Recurrence in Non-Muscle-Invasive Bladder Cancer after Bacillus Calmette-Guérin Therapy

Seung-Woo Baek^{1,#}, Young Joon Byun^{2,#}, Xuan-Mei Piao², Jee-Woo Seo^{1,3}, Il-San Jeong^{1,3}, Otabek Kamarov², Seongmin Moon⁴, Kyeong Kim^{2,4}, Ho Won Kang⁵, Won Tae Kim^{2,4}, Yung Hyun Choi⁶, Sung-Kwon Moon⁷, Wun-Jae Kim⁸, Young Deuk Choi⁹, Seon-Young Kim^{1,3}, Seon-Kyu Kim^{1,3,*}, Seok Joong Yun^{2,4,10,*}

¹Genomic Medicine Research Center, Korea Research Institute of Bioscience and Biotechnology, Daejeon, Republic of Korea

²Department of Urology, Chungbuk National University College of Medicine, Cheongju, Republic of Korea

³Department of Bioscience, University of Science and Technology, Daejeon, Republic of Korea

⁴Department of Urology, Chungbuk National University Hospital, Cheongju, Republic of Korea

⁵Department of Urology, Hankook General Hospital, Cheongju, Republic of Korea

⁶Department of Biochemistry, Dong-eui University College of Korean Medicine, Busan, Republic of Korea

⁷Department of Food and Nutrition, Chung-Ang University, Ansong, Republic of Korea

⁸Institute of Urotech, Cheongju, Republic of Korea

⁹Department of Urology and Urological Science Institute, Yonsei University College of Medicine, Seoul, Republic of Korea

¹⁰Cancer Research Center, Chungbuk National University, Republic of Korea

[#]These authors contributed equally to this work

*Corresponding authors: Seok Joong Yun, MD, PhD, Department of Urology, Chungbuk National University College of Medicine, Chungdae-ro 1, Seowon-Gu, Cheongju, Chungbuk 28644, Republic of Korea (sjyun@chungbuk.ac.kr). Seon-Kyu Kim, PhD, Genomic Medicine Research Center, Korea Research Institute of Bioscience and Biotechnology, 125 Gwahak-ro, Yuseong-gu, Daejeon 34141, Republic of Korea. (seonkyu@kribb.re.kr)

ARTICLE IN PRESS

Abstract

Background: High-frequency recurrence (HfR, ≥ 2 recurrences) in non-muscle-invasive bladder cancer (NMIBC) poses a significant clinical burden. Current risk models, such as the European Organization for Research and Treatment of Cancer (EORTC), the European Association of Urology (EAU), and the UROMOL classification offer limited predictive accuracy for identifying patients at risk for frequent recurrence despite appropriate treatment.

Methods: A 75-gene high-frequency recurrence index (HfRI) was constructed by selecting recurrence-associated genes using differential expression and Cox regression analyses. The HfRI was computed as a weighted sum of normalized gene expression values. The model was trained on a discovery cohort and validated in multiple cohorts (n=1,379) using machine learning approaches. Clinical relevance was assessed using recurrence-free survival (RFS) and Cox models, and predictive performance was compared with that of the EORTC, EAU, and UROMOL classifications using the area under the curve (AUC) and the concordance index (c-index).

Results: The HfRI robustly stratified patients into high-and low-risk groups across six independent NMIBC cohorts. Patients classified as HfRI-high had a significantly greater likelihood of experiencing ≥ 2 recurrences (χ^2 , $p=0.001$) and showed markedly reduced RFS (log-rank test, $p<0.001$). The adverse prognostic effect of the HfRI persisted even among patients treated with BCG therapy (log-rank test, $p=0.02$). Multivariate analysis revealed that the HfRI was an independent predictor of HfR (HR=2.82, 95% CI=1.89–4.20; $p<0.001$). Compared with established clinical risk classifiers, the HfRI demonstrated superior predictive performance (AUC=0.736; c-index=0.673) in terms of the EORTC (AUC=0.594), EAU (AUC=0.557) risk groups, and UROMOL2021 (AUC=0.596) classification. Pathway analysis revealed that HfRI-high tumors were characterized by upregulation of cell cycle progression and DNA replication pathways, accompanied by suppression of immune signaling pathways. These biological features provide a mechanistic explanation for the reduced responsiveness to intravesical BCG therapy, underscoring the role of HfRI not only as a predictor of recurrence risk but also as a biomarker capable of identifying patients unlikely to benefit from

standard BCG treatment.

Conclusions: HfRI represents a robust, transcriptome-based tool for predicting frequent recurrence in NMIBC patients. The HfRI supports earlier identification of patients at risk of high-frequency recurrence, thereby supporting personalized treatment strategies.

Keywords

NMIBC, Frequent recurrence, High-frequency recurrence index, BCG therapy

ARTICLE IN PRESS

Background

Bladder cancer (BC) is among the top ten most common cancers globally [1], with 70% of cases diagnosed as non-muscle invasive bladder cancer (NMIBC) [2]. While NMIBC has a relatively favorable prognosis compared with muscle-invasive BC, its clinical management is complicated by high recurrence rates, which have been reported in as high as 70% of cases [3]. Even after appropriate transurethral resection (TUR) and intravesical *Bacillus Calmette-Guérin* (BCG) therapy, approximately 20% of patients experience frequent recurrences (defined as two or more episodes) [4, 5], necessitating repeated procedures that negatively affect quality of life and increase health care burden. NMIBC includes tumors confined to the mucosa (stage Ta), tumors invading the subepithelial connective tissue (T1), and flat carcinoma in situ (Tis), all of which lack invasion into the muscle layer. These early-stage tumors are typically treated with TUR followed by intravesical therapy [4]. However, disease behavior varies widely even within these noninvasive stages, and recurrence risk is influenced by factors such as tumor size, multiplicity, prior recurrence, and histologic grade. Accurate risk stratification is therefore essential for optimizing treatment intensity and follow-up strategies.

To guide treatment decisions for NMIBC patients who experience recurrence, various clinical and molecular risk stratification models have been proposed, including the European Organization for Research and Treatment of Cancer (EORTC) and the European Association of Urology (EAU) scores, as well as transcriptomic classifications such as UROMOL2021 [4, 6, 7]. However, these models were not specifically designed to predict high-frequency recurrence (HfR), and their accuracy in this context remains low. Moreover, most molecular studies to date have focused on progression to muscle-invasive disease [7-13], whereas relatively few have addressed the prediction of frequent recurrence, which is a clinically important yet underexplored outcome in NMIBC. A more robust and biologically grounded predictor is therefore needed to guide individualized management, particularly for patients who may not benefit from repeated BCG therapy.

We have previously conducted a range of studies to predict clinical events such as progression and recurrence in NMIBC patients, as well as BCG treatment responsiveness [10, 13-15].

Building upon this experience, we sought to address one of the most critical unmet clinical needs: predicting high-frequency recurrence and its association with the response to BCG treatment, an area that has not been systematically explored. We used transcriptomic data from both newly generated and publicly available cohorts, focusing on patients with ≥ 2 recurrences despite having similar clinical risk scores. The high-frequency recurrence index (HfRI) was developed through Cox regression analysis of recurrence-associated genes and validated using machine learning across six independent cohorts. In addition to its predictive performance, HfRI-high tumors displayed transcriptomic patterns linked to cell proliferation and immune suppression, providing a biological context that may help refine patient stratification and guide future investigations. These findings support the clinical utility of the HfRI as a robust predictor of frequent recurrence and as a tool for precision-guided management of NMIBC.

ARTICLE IN PRESS

Methods

Terminology clarification

To avoid ambiguity, we distinguished between the two classification schemes used in this study: (1) Outcome-defined groups (HfR-NMIBC, ≥ 2 recurrences; NR-NMIBC, no recurrence) were used only in the discovery cohort to identify recurrence-associated genes and construct the 75-gene signature. (2) Model-based classifications (HfRI-high/HfRI-low) were applied in all downstream analyses after the establishment of the HfRI. This distinction is maintained throughout the manuscript to ensure consistent terminology.

Tissue samples from the discovery and validation cohorts

Patient-derived tumor tissues were collected to generate transcriptome data for patients with high-frequency recurrence of NMIBC (GSE294703, n=45) [16]. Additional transcriptome data for NMIBC (GSE294705, n=94) were generated in this study for validation [17]. These patient-derived tissues were collected through appropriate procedures after surgery at Chungbuk National University Hospital, Cheongju, Republic of Korea. All transcriptome data used in this study included patient recurrence-free survival (RFS), follow-up period, and frequency of recurrence. The baseline characteristics of the clinical cohorts are described in **Table 1** and Additional file 1: **Table S1**. The tumor tissue was macrodissected within 15 minutes after surgical resection. To ensure representativeness, adjacent tissue from fresh frozen sections of transurethral resection specimens was analyzed. The samples were then stored in liquid nitrogen at -80°C until use. Sample collection and analysis were approved by the Chungbuk National University Institutional Review Board and by the Chungbuk National University Hospital (IRB approval number: CBNU-202309-BR-0201 and GR2010-12-010, respectively), and all patients provided informed consent. High-frequency recurrence is defined as 2 or more recurrences. The workflow of this study and the information on the clinical cohorts utilized at each step are described in Additional file 2: **Fig. S1**.

RNA extraction, RNA-seq experiments, and data processing

Total RNA was extracted from tissues using TRIzol reagent (Invitrogen, Carlsbad, CA) and stored at -80 °C. RNA quality and integrity were assessed using the RNA integrity number and the DV200 metric, which were determined with an RNA 6000 Nano Kit and an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA). Samples with a RIN >8 were considered good quality and were selected for downstream analysis. The sequencing library was prepared using the TruSeq RNA Sample Preparation Kit v2 (Illumina, San Diego, CA). Briefly, mRNA was purified from total RNA using poly-T oligo-attached magnetic beads, fragmented, and converted into cDNA. Adapters were ligated to the cDNA, and the fragments were amplified by polymerase chain reaction. Paired-end sequencing reads (2 × 150 bp) were obtained using the HiSeq 2500 platform (Illumina, San Diego, CA). The reference genome sequence data for *Homo sapiens* were obtained from the NCBI Genome database (assembly ID: GRCh38), and reference genome indexing and read mapping of tissue samples were performed using STAR software (version 2.7.3a, RRID:SCR_004463) [18].

Definition of High-Frequency Recurrence

HfR was operationally defined as two or more documented tumor recurrences (≥ 2) following the initial diagnosis of NMIBC. This definition was consistently applied across all cohorts that provided recurrence frequency information, including both in-house and publicly available datasets. Data on recurrence timing, such as the intervals between recurrences or recurrence rates per year, were not uniformly available across several public datasets. Consequently, temporal dynamics could not be incorporated into the definition. To ensure consistency across cohorts with heterogeneous reporting formats, the recurrence count was selected as the primary endpoint.

BCG treatment

Patients with intermediate- and high-risk NMIBC underwent an induction course of intravesical BCG consisting of six weekly instillations following TUR. For each instillation, one vial of BCG (12.5 mg per vial for OncoTice[®]) was reconstituted in saline and instilled into the bladder. Cystoscopy was performed three months after the completion of BCG induction therapy and every three months thereafter. If there was no recurrence or progression for two years, cystoscopy was subsequently

performed every six months. If a lesion suggestive of recurrence was detected during cystoscopy, a TUR procedure was performed. In cases where no visible lesions were detected, patients received a maintenance regimen consisting of three weekly BCG instillations at 3, 6, 12, and 18 months after surgery.

Public datasets of NMIBC patients

We obtained clinical information and gene expression data from NMIBC patients from public databases. The raw data from the UROMOL consortium were downloaded from the European Genome-Phenome Archive (EGA) under accession number EGAS0001004693 (UROMOL2021) [7, 19]. Gene expression data for the bladder cancer microarray study were downloaded from the Gene Expression Omnibus (GEO, RRID:SCR_005012) database from the National Center for Biotechnology Information (NCBI) under accession numbers GSE5479 [12, 20], GSE13507 [13, 14, 21], GSE120736 [11, 22], and GSE163209 [8, 23]. The baseline characteristics of patients whose data were collected from public repositories are provided in Additional file 1: **Table S1**, alongside those of GSE294705 [17].

Data preprocessing for transcriptomic profiling

All gene expression data were separately log₂ transformed and quantile normalized. For the RNA-seq data, the read counts per million (CPM) fragments mapped of each sample were used to estimate the expression level of each gene. Genes with an average CPM value < 1 for the entire sample were removed.

Development of the HfRI

To develop the HfRI, 700 differentially expressed genes were identified between the two groups in GSE294703 by applying a 1.5-fold change threshold, $p < 0.05$ (two-sample t test), and a false discovery rate (FDR) adjustment for multiple comparisons [16]. We then used Cox regression analysis of these genes in three independent NMIBC cohorts to select 75 genes that maintained consistent prognostic associations. This gene was defined as a HfR-related gene. HfRI was developed to quantify the transcriptomic signature associated with frequent recurrence in NMIBC. For each patient, gene

expression values were first standardized (z-score). The HfRI was then calculated as the sum of the standardized expression levels of 75 genes, each multiplied by its corresponding Cox regression coefficient. These coefficients were derived as the average from three independent NMIBC cohorts (UROMOL2021, GSE13507, and GSE120736), considering only genes that consistently showed positive (hazard ratio [HR] > 1) or negative (HR < 1) associations with high-frequency recurrence. To determine the optimal cutoff for the HfRI, we used the nearest-neighbor estimation method to generate time-dependent receiver operating characteristic (ROC) curves for RFS.

Development of a machine learning-based prediction model

To verify the prognostic relevance of the subgroups derived from the HfRI, machine learning algorithms were used to generate predictive models. These algorithms include random forest (R package, randomForest, version 4.7-1.1), support vector machine (R package, e1071, version 1.7-16), and XGBoost (R package, xgboost version 1.7.8.1). Among them, the model with the highest balanced accuracy was selected. In the UROMOL2021 cohort, the ratio of training data to testing data was 8:2. The performance was improved by modifying hyperparameters appropriate for the algorithm used through grid search. The gene expression data used to construct the predictive models were preprocessed through normalization and standardization.

Biological enrichment pathway and other statistical analyses

To explore the diversity of biological processes across subgroups, we used gene set variation analysis (GSVA) [24]. For this purpose, we utilized gene sets from the Molecular Signatures Database (MSigDB), which includes both HALLMARK and the Kyoto Encyclopedia of Genes and Genomes (KEGG, RRID:SCR_012773). We normalized the GSVA scores and evaluated differences in expression between the two groups using a *p* value threshold of < 0.05, with a focus on identifying statistically significant alterations. In addition, gene set enrichment analysis (GSEA) was performed with significance criteria (FDR<0.25). Kaplan–Meier plots were used to estimate associations between subgroups and survival outcomes using the log-rank test. Clinical variables, such as clinical information that were distributed differently across subgroups were estimated by chi-square or

Fisher's exact tests, which are appropriate for the characteristics of the data. ROC curve analysis was used to calculate the area under the curve (AUC). The predictive HfRI values were assessed using Harrell's concordance index (c-index). In addition, statistical significance was evaluated with DeLong's test. All statistical analyses were performed in the R version 4.2.1 language environment (<https://www.r-project.org>, RRID:SCR_000432).

Data availability

The RNA-seq datasets generated in this study are available in the Gene Expression Omnibus (GEO) repository under accession numbers GSE294703[16] and GSE294705[17]. The public datasets analyzed in this study include the GEO accession numbers GSE5479[12, 20], GSE13507[13, 14, 21], GSE120736[11, 22], and GSE163209[8, 23] and the European Genome-Phenome Archive (EGA) accession number EGAS0001004693 (UROMOL2021; FASTQ files and clinical information)[7, 19]. R functions for calculating the HfRI and the trained machine-learning prediction models are available at <https://github.com/baeksw33/HighFrequencyRecurIndexNMIBC>.

Results

Molecular characterization of HfR in BCG-treated NMIBC

Building upon our previous studies predicting progression, recurrence, and BCG responsiveness in NMIBC [10, 13-15], we undertook a study to address one of the most pressing unmet needs in clinical practice: the prediction of high-frequency recurrence. A summary of our study workflow is shown in Additional file 2: **Fig. S1**. We analyzed transcriptome data from 45 NMIBC patients treated at Chungbuk National University Hospital, all of whom underwent TUR and BCG therapy. Among them, 22 patients experienced no recurrence after BCG treatment, while 23 experienced frequent recurrence. Importantly, EORTC risk scores did not significantly differ between the two groups. Patients without recurrence were classified as nonrecurrent NMIBC (NR-NMIBC), and those with two or more recurrences were classified as HfR-NMIBC patients (**Table 1**).

The clinicopathological variables did not significantly differ between these subgroups (Additional file 2: **Fig. S2A** and **Table 1**). To reveal the biological pathways associated with recurrence, we performed GSEA using HALLMARK and KEGG pathway data from MSigDB. HfRI-high patient samples exhibited upregulation of E2F targets, mitotic spindle, and glycolysis, and downregulation of the interferon alpha and gamma response pathways (Additional file 2: **Fig. S2B**). Key genes involved in these pathways, including *CENPJ*, *BUB1*, and *MKI67* as well as *GPC3*, have previously been linked to recurrence in other cancers [25-29]. Conversely, the expression of immune-related genes such as *CXCL9* and *CXCL10* [30, 31] was suppressed in HfRI-high patients (Additional file 2: **Fig. S2A**). We identified 700 genes that were differentially expressed between HfRI-high patients and NR-NMIBC patients (Additional file 2: **Fig. S2A**). To assess prognostic relevance, we performed hierarchical clustering of the UROMOL2021 cohort using these genes. Compared with the NR-NMIBC subgroup, the HfR-NMIBC subgroup had significantly shorter RFS (log-rank test, $p=0.03$; Additional file 2: **Figs. S3A, S3B**), with an even more pronounced difference among patients who experienced two or more recurrences (log-rank test, $p=0.002$; Additional file 2: **Fig. S3C**). The frequency of recurrence was also higher in the HfRI-high patients (χ^2 test, $p=0.007$; Additional file 2:

Fig. S3D), supporting the relevance of the gene signature for stratifying NMIBC by high-frequency recurrence.

Construction of the HfRI for predicting HfR

To facilitate clinical implementation while preserving specificity for HfR, we refined the original 700 HfR-focused genes into a streamlined 75-gene signature that retains predictive power but is feasible for diagnostic assay development (Additional file 1: **Table S2**). The coefficients derived from Cox regression analysis were used as weights to design the HfRI. We applied the HfRI to the UROMOL2021 cohort and stratified patients into HfRI-high and HfRI-low subgroups. To determine the optimal cutoff value, we used high-frequency recurrence status (≥ 2 recurrences) as the binary outcome, and an ROC curve was constructed. Comparisons with molecular features from UROMOL2021 revealed that the characteristics of the HfRI-high patients were consistent with those of our previous GSEA (Additional file 2: **Fig. S2**), indicating the activation of pathways related to the late-stage cell cycle and DNA replication. Additionally, uroplakin and pan-fibroblast TGF β signatures were enriched in HfRI-high patients. In contrast, the HfRI-low subgroup displayed overall immune activation, such as total immune score, CD274 expression, and the interferon-gamma signature (Additional file 2: **Fig. S4A**).

Using the log-rank test, we confirmed that compared with the HfRI-low group, the HfRI-high group experienced significantly shorter RFS (log-rank test, $p < 0.001$; **Fig. 1A**). This distinction remained when patients with only a single recurrence were excluded (log-rank test, $p < 0.001$; **Fig. 1B**). Similarly, the proportion of patients with ≥ 2 recurrences was much greater in the HfRI-high group (χ^2 test, $p < 0.001$; **Fig. 1C**). We observed the same trend in the smaller subset of patients who had received prior BCG therapy, but the limited sample size precluded statistical confirmation (Additional file 2: **Fig. S4B**). Multivariate Cox proportional hazard regression analysis of UROMOL2021 identified the HfRI as an independent risk factor for RFS ≥ 2 recurrences (hazard ratio [HR] = 3.23; 95% confidence interval [CI]=1.96–5.34; Additional file 1: **Table S3**). HfRI demonstrated good predictive performance, with an AUC of 0.736 and a c-index of 0.673 (**Fig. 1D**), outperforming conventional clinicopathologic and molecular classifications, including stage, grade,

EORTC and EAU risk score, and Class 2a subtype of UROMOL (AUC range: 0.524–0.596; Additional file 2: **Fig. S4C**). To further examine whether integrating clinical variables could enhance predictive performance, we constructed combined models incorporating the HfRI with EORTC and EAU risk scores. Compared with HfRI alone, both combined models modestly improved the prediction accuracy: the model with HfRI alone; HfRI + EORTC + EAU achieved an AUC of 0.750 (c-index=0.696) (**Fig. 1D**). Although these improvements were incremental, they suggest that integrating selective clinical indicators with the HfRI may provide additional prognostic refinement. Nevertheless, HfRI alone remained the strongest individual predictor, achieving the most favorable separation of the RFS curve (Additional file 2: **Fig. S4D–G**). We next assessed how HfRI aligns with existing molecular and clinical subtype systems. HfRI-high patients were significantly enriched in the high-risk categories of both the EORTC and the EAU clinical risk groups (Additional file 2: **Fig. S4H and 4I**). At the transcriptomic level, HfRI-high tumors were strongly enriched for the proliferative UROMOL Class 2a subtype and the Leeds E3 subtype, both of which represent aggressive molecular backgrounds (Additional file 2: **Fig. S4J and 4K**). These subtype distributions further support the association of HfRI-high tumors with aggressive and recurrence-prone tumors.

Validation of the HfRI in independent NMIBC cohorts

After confirming the clinical utility of the HfRI in the UROMOL2021 cohort, we extended our investigation to characterize these subgroups across additional independent NMIBC cohorts. Using the 75-gene-based HfRI, we applied three machine learning algorithms to classify patients in these cohorts (Additional file 1: **Table S4**). Among the models tested, the random forest algorithm demonstrated the best performance, achieving an accuracy of 93% in the UROMOL2021 cohort, and was selected as the final HfRI-subtype predictor for further analysis (Additional file 2: **Fig. S5**).

When we applied the model to the GSE13507 cohort, the 75-gene HfRI clearly differed in terms of expression profiles (Additional file 2: **Fig. S6A**) and separated RFS curves (log-rank test, $p = 0.02$; Additional file 2: **Fig. S6B**). This separation persisted after single-recurrence cases were excluded (log-rank test, $p = 0.01$; Additional file 2: **Fig. S6C**). Consistently, ≥ 2 recurrences were more common in the HfRI-high group (χ^2 test, $p = 0.04$; Additional file 2: **Fig. S6D**). The AUC (0.676)

and c-index (0.642) values indicated moderate predictive accuracy (Additional file 2: **Fig. S6E**). For completeness, we explored a smaller subset of patients who had received prior BCG therapy; the same trend was observed, but the numbers were insufficient to reach statistical significance (Additional file 2: **Fig. S6F**). In the GSE120736 cohort, similar expression patterns for the 75 HfR-related genes were observed (Additional file 2: **Fig. S7A**), with the HfRI-high patients tending toward a higher recurrence rate (log-rank test, $p = 0.09$; Additional file 2: **Fig. S7B**). When patients with only a single recurrence were excluded, the separation became significant (log-rank test, $p = 0.008$; Additional file 2: **Fig. S7C**), and the proportion of patients with ≥ 2 recurrences was markedly greater in the HfRI-high group (χ^2 test, $p < 0.001$; Additional file 2: **Fig. S7D**). The HfRI demonstrated good predictive performance, with an AUC of 0.728 and a c-index of 0.673 (Additional file 2: **Fig. S7E**). We applied the HfRI classification to a newly generated validation cohort (GSE294705), in which similar expression patterns for the 75 HfR-related genes were observed (Additional file 2: **Fig. S8A**). In this cohort, compared with the HfRI-low subgroup, the HfRI-high subgroup had a significantly higher recurrence rate (log-rank test, $p = 0.03$; Additional file 2: **Fig. S8B**). After patients with only a single recurrence were excluded, HfRI effectively stratified patients with respect to disease recurrence (log-rank test, $p = 0.01$; Additional file 2: **Fig. S8C**), with a significantly greater proportion of ≥ 2 recurrence cases observed in the HfRI-high patients (χ^2 test, $p = 0.01$; Additional file 2: **Fig. S8D**). The HfRI exhibited modest predictive power, yielding an AUC of 0.613 and a concordance index of 0.562 (Additional file 2: **Fig. S8E**). Consistent results emerged in the GSE163209 and GSE5479 cohorts: the 75-gene HfRI pattern was reproduced, and the HfRI-high group had worse RFS (log-rank tests, $p = 0.04$ and $p = 0.07$, respectively; Additional file 2: **Figs. S9A, S9B, S9D, and S9E**). The same trend was observed among the BCG-treated patients, with higher recurrence rates in the HfRI-high patients (log-rank tests, $p = 0.11$ and $p = 0.03$, respectively; Additional file 2: **Figs. S9C and S9F**). In multivariate Cox models, the HfRI remained an independent predictor of RFS in both the UROMOL2021 cohort (HR = 1.58; 95% CI 1.21–2.05) and the pooled cohort (HR = 1.65; 95% CI 1.34–2.02) (Additional file 1: **Table S5**).

Multicohort integration confirms HfRI as a robust predictor of HfR and limited BCG benefit

All cohorts reporting recurrence frequency (UROMOL2021, GSE13507, GSE120736, and GSE294705) were combined into a single total cohort. After patients with only one recurrence were excluded, the RFS curves were clearly stratified (log-rank test, $p < 0.001$; **Fig. 2A**). Because current EAU guidelines recommend BCG for intermediate- to high-risk NMIBC, we next evaluated 148 BCG-treated patients within this total cohort (single-recurrence cases excluded). Patients classified as HfRI-high derived significantly less benefit from BCG and exhibited markedly shorter RFS than HfRI-low patients did (log-rank test, $p = 0.02$; **Fig. 2B**). These findings indicate that BCG therapy is substantially more effective in patients classified as HfRI-low and that alternative or intensified therapeutic strategies may be warranted for the HfRI-defined high-risk group. Consistent with these survival patterns, HfRI-high patients had a significantly greater percentage of patients with ≥ 2 recurrences (χ^2 test, $p = 0.001$; **Fig. 2C**). Multivariate Cox analysis revealed that the HfRI remained an independent predictor of RFS (univariate HR = 3.29, 95% CI 2.36–4.58; multivariate HR = 2.82, 95% CI 1.89–4.20; **Table 2**). Overall, the HfRI demonstrated acceptable discriminative performance, with an AUC of 0.701 and a c-index of 0.653 (**Fig. 2D**). To further examine whether integrating clinical variables could enhance predictive performance, we constructed combined models incorporating the HfRI with EORTC and EAU risk scores. Compared with HfRI alone, both combined models modestly improved prediction accuracy: the model with HfRI alone: HfRI + EORTC + EAU achieved an AUC of 0.708 (c-index=0.661) in the total cohort (**Fig. 2D**).

In addition to the reduced clinical benefit from BCG, HfRI-high patients displayed a distinct molecular phenotype characterized by activation of the pan-fibroblast TGF β signature and suppression of immune activity, including a reduced total immune score and decreased CD274 expression (Additional file 2: **Fig. S10**). These features are consistent with an immune-excluded, stromal-driven tumor microenvironment and further support the need to explore alternative or immunomodulatory strategies for patients with high HfRI scores.

Taken together, these data validate the HfRI in multiple independent cohorts, highlight its robustness in identifying patients at risk for HfR, and suggest its potential utility in guiding treatment decisions, particularly in identifying individuals unlikely to benefit from standard BCG treatment.

Discussion

In this study we introduce the high-frequency recurrence index (HfRI), a transcriptome-based tool designed to predict high-frequency recurrence (HfR; ≥ 2 recurrences) in NMIBC patients. Despite BCG therapy, recurrence affects approximately 50–70% of patients, necessitating repeated cystoscopies, TURs, and intravesical instillations [2, 3, 32]. These interventions, while essential, impose considerable physical, psychological, and financial burdens [33]. Moreover, single recurrences may result from undetected lesions or technical variability rather than tumor-intrinsic biology. Therefore, biomarkers that capture the underlying biological risk of sustained recurrence are urgently needed. The HfRI was developed to address this gap and offers several key advantages.

First, we validated the prognostic relevance of the HfRI across both in-house and public NMIBC cohorts. Patients classified as HfRI-high consistently had shorter RFS and more frequent recurrences, even after single-recurrence cases were excluded, supporting its biological relevance. Second, compared with established tools such as the EORTC, EAU, and UROMOL2021 classifiers, the HfRI demonstrated relatively stronger predictive performance for identifying patients at high risk of HfR. Third, HfRI-high patients displayed features associated with reduced BCG responsiveness. Activation of the pan-fibroblast TGF β pathway and suppression of immune-related signatures, such as decreased total immune score and CD274 expression, suggest an immune-excluded, stromal-driven microenvironment that may limit the efficacy of BCG [31, 34, 35]. The biological plausibility of HfRI is further supported by the presence of key genes with well-established roles in aggressive tumor biology. For example, MKI67 and BUB1 are widely recognized markers of proliferative activity and treatment resistance across multiple solid tumors, including breast cancer [27, 28]. In addition, UPK3B reflects urothelial differentiation programs and is associated with recurrence and disease progression in mesothelial and epithelial tumor contexts [36].

In addition to its prognostic performance, the HfRI showed distinct relationships with established molecular subtypes. HfRI-high patients were markedly enriched in the high-risk group according to the EORTC and EAU clinical risk models and aligned strongly with the proliferative UROMOL Class

2a and Leeds E3 molecular subtypes. These associations indicate that HfRI captures tumors positioned at the aggressive end of existing clinical and transcriptomic classifications. Compared with Class 2a and E3 patients, patients with high HfRI levels showed additional features, such as a reduced total immune score and activation of stromal TGF β signaling, which are biologically relevant pathways captured by HfRI.

Although the HfRI was cross-validated in six independent cohorts, the retrospective design and cohort heterogeneity may introduce selection bias and limit the generalizability of the findings. The absence of a prospective validation cohort remains a key limitation of this study, as retrospective analyses cannot fully substitute for real-world or forward-looking clinical evaluation. To partially address this, we also evaluated the HfRI in an independent, consecutively accrued institutional validation cohort (GSE294705) comprising intermediate- or high-risk NMIBC patients, almost all of whom received guideline-based intravesical BCG, in which the HfRI showed a similar association with high-frequency recurrence. Future prospective studies will be essential to confirm the clinical utility, reproducibility, and implementation feasibility of the HfRI. In addition, the number of BCG-treated patients with ≥ 2 recurrences in several cohorts was relatively small, reducing the statistical power for subgroup analyses and potentially leading to unstable risk estimates. Nevertheless, the reproducibility of the HfRI across the six datasets indicates that it captures a stable biological signal and provides a solid foundation for future prospective refinement.

Beyond its prognostic utility, an important limitation of this study is that we did not experimentally evaluate alternative therapeutic strategies for the HfRI-high subgroup, despite demonstrating that these patients respond poorly to BCG. However, the molecular phenotype of HfRI-high tumors, characterized by immune exclusion and stromal TGF β pathway activation, offers a biologically coherent explanation for reduced BCG responsiveness and provides direction for future therapeutic exploration. Previous studies have shown that stromal TGF β signaling promotes T-cell exclusion and resistance to immune-based therapies [34], whereas combined PD-L1/TGF β blockade can restore cytotoxic T-cell infiltration in immune-excluded tumors [37]. These observations suggest that therapies targeting the stromal TGF β axis or enhancing immune infiltration may represent rational

alternatives for HfRI-high NMIBC and warrant investigation in further studies.

HfRI has the potential to be incorporated into existing NMIBC workflows following TURBT and pathology review. Gene expression profiling can be performed using RNA-seq, qRT-PCR, or NanoString platforms, all of which are routinely available in clinical molecular laboratories. Because the HfRI relies on a fixed 75-gene panel and is computed from normalized gene-level expression values, it is fully compatible with all three platforms and does not require any additional clinical procedures. Once expression data are generated, HfRI scores can be calculated immediately, enabling timely and workflow-aligned risk stratification. At the same time, although the current 75-gene HfRI provides robust predictive performance, it should be regarded as a research-grade signature at this stage. For routine clinical deployment, further gene reduction and assay optimization will likely be necessary to improve feasibility, cost-effectiveness, and scalability while preserving predictive accuracy. Taken together, this implementation framework supports the integration of HfRI into routine practice and suggests that HfRI-low patients may benefit from standard BCG-based guidelines, whereas HfRI-high patients may require intensified or alternative therapeutic strategies (Additional file 2: **Fig. S11**).

5. Conclusions

The HfRI represents a promising prognostic tool for identifying NMIBC patients at high risk of high-frequency recurrence, enabling early intervention and guiding tailored treatment strategies. Expanding the clinical use of HfRI will require further prospective validation and exploration of personalized treatment approaches. Future studies could evaluate immunotherapies, stromal-targeted therapies, or combination strategies for HfRI-high patients. Additionally, efforts to standardize and clinically translate the HfRI, including platform adaptation and cost-effectiveness assessments, will facilitate its integration into routine practice.

Table 1. Baseline characteristics of the discovery cohort (GSE294703)

Variables	HfR-NMIBC	NR-NMIBC	p-value
Total	23	22	
Age (mean (SD))	65.35 (10.68)	69.36 (9.10)	0.182
Gender (%)			
Female	3 (13.0)	6 (27.3)	0.412
Male	20 (87.0)	16 (72.7)	
Stage (%)			
T1	12 (52.2)	15 (68.2)	0.429
Ta	11 (47.8)	7 (31.8)	
Grade (%)			
High	10 (58.8)	14 (73.7)	0.555
Low	7 (41.2)	5 (26.3)	
Tumor.Num. (%)			
≥8	8 (34.8)	10 (45.5)	0.640
2~7	7 (30.4)	7 (31.8)	
1	8 (34.8)	5 (22.7)	
Tumor.Size (%)			
<1cm	4 (17.4)	0 (0.0)	0.047
<3cm	6 (26.1)	3 (13.6)	
≥3cm	13 (56.5)	19 (86.4)	
CIS (%)			
CIS	0 (0.0)	1 (4.5)	0.982
No CIS	23 (100.0)	21 (95.5)	
BCG (%)			
BCG treatment	23 (100.0)	22 (100.0)	NA
EORTC-recr score, n (%)			
High (6 >)	12	15	0.410
Low (6 ≤)	11	7	
Recurrence event (%)			
No recurrence	0 (0.0)	22 (100.0)	<0.001
Recurrence	23 (100.0)	0 (0.0)	
Frequency of recurrence (min-max (median))	2-7 (3.0)	0 (0.0)	<0.001
Follow up (mean (SD))	85.42 (49.87)	53.60 (49.20)	0.037
Abbreviation: CIS, Carcinoma in situ; BCG, Bacillus Calmette-Guérin; HfR, High-frequency recurrence; NR, No recurrence; EORTC, European Organisation for Research and Treatment of Cancer; SD, Standard Deviation.			

Table 2. Univariate and multivariate Cox proportional analysis of high-frequency of recurrence (≥ 2) of patients with total cohort

	Variable		Univariate					Multivariate			
			n	HR	ciLower	ciUpper	p-value	HR	ciLower	ciUpper	p-value
Total cohort	Gender	Male (Ref.) vs. female	485	2.35	1.70	3.24	< 0.001	1.56	0.99	2.44	0.055
	Age	≤ 65 (Ref.) vs. > 65	423	1.66	1.15	2.39	0.007	1.26	0.84	1.89	0.272
	Stage	Ta (Ref.) vs. T1	479	1.09	0.79	1.50	0.606	0.70	0.42	1.16	0.167
	Grade	Low (Ref.) vs. High	481	1.13	0.82	1.56	0.449	1.06	0.69	1.64	0.784
	BCG treatment	No (Ref.) vs. Yes	370	1.42	0.96	2.10	0.083	1.53	0.94	2.48	0.084
	HfRI	Low (Ref.) vs. High	487	3.29	2.36	4.58	< 0.001	2.82	1.89	4.20	< 0.001

ARTICLE IN PRESS

Abbreviations

BC, Bladder cancer

NMIBC, Non-muscle invasive bladder cancer

TUR, Transurethral resection

BCG, Bacillus Calmette-Guérin

EORTC, European Organization for Research and Treatment of Cancer

EAU, European Association of Urology

HfR, High-frequency Recurrence

HfRI, High-frequency Recurrence Index

RFS, Recurrence-Free Survival

EGA, European Genome-phenome Archive

GEO, Gene Expression Omnibus

NCBI, National Center for Biotechnology Information

CPM, Counts Per Million fragments mapped

FDR, False Discovery Rate

ROC, Receiver Operating Characteristic

GSVA, Gene Set Variation Analysis

MSigDB, Molecular Signatures Database

KEGG, Kyoto Encyclopedia of Genes and Genomes

GSEA, Gene Set Enrichment Analysis

AUC, Area under the curve

c-index, Harrell's concordance index

NR-NMIBC, Non-Recurrence NMIBC

HR, Hazard ratio

CI, Confidence interval

ARTICLE IN PRESS

Supplementary Information

Additional file 1. Table S1-S5. Table S1-[Baseline characteristics of the independent NMIBC clinical cohorts]. Table S2-[Cox regression analysis of genes comprising the High-frequency Recurrence Index]. Table S3-[Univariate and multivariate Cox proportional analysis of high-frequency of recurrence (≥ 2) of patients with UROMOL2021]. Table S4-[Performance metrics of machine learning models in predicting HfRI in the UROMOL2021 cohort]. Table S5-[Univariate and multivariate Cox proportional analysis of recurrence-free survival of all patients with NMIBC].

Additional file 2. Figure S1-S11. Fig S1-[Workflow and clinical cohorts utilized in this study]. Fig S2-[Molecular characteristics of high-frequency recurrence (HfR) versus non-recurrence (NR) NMIBC in the discovery cohort (GSE294703)]. Fig S3-[Prognostic relevance between NR-NMIBC and HfR-NMIBC in UROMOL2021 cohort]. Fig S4-[Comparison of subgroup characteristics and prognostic relevance in NMIBC patients by high-frequency recurrence index (HfRI) in UROMOL2021 cohort]. Fig S5-[Development and validation of a machine learning-based classifier for high-frequency recurrence in NMIBC]. Fig S6-[Comparison of subgroup characteristics and prognostic relevance in NMIBC patients by high-frequency recurrence index (HfRI) in GSE13507 cohort]. Fig S7-[Comparison of subgroup characteristics and prognostic relevance in NMIBC patients by high-frequency recurrence index (HfRI) in GSE120736 cohort]. Fig S8-[Comparison of subgroup characteristics and prognostic relevance in NMIBC patients by high-frequency recurrence index (HfRI) in validation cohort (GSE294705)]. Fig S9-[Comparison of subgroup characteristics and prognostic relevance in NMIBC patients by high-frequency recurrence index (HfRI) in GSE163209 and GSE5479 cohort]. Fig S10-[Comparison of pan-fibroblast TGF β signaling, total immune score, and CD274 (PD-L1) expression by HfRI]. Fig S11-[Clinical workflow and decision strategy for implementing the High-Frequency Recurrence Index (HfRI) in NMIBC].

Declarations

Acknowledgments

Not applicable.

Consent for publication

Not applicable.

Funding

This research was supported by grant from the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2022R1A6A3A01087141, and RS-2023-00245919) and by the Ministry of Science and ICT (MSIT) (No. RS-2025-19922981, RS-2024-00342111, NRF-2021M3H9A2096953, and NRF-2022R1A2C2008480) and by the BK21 FOUR of the NRF funded by the Ministry of Education (No. 2120240615350) and by Chungbuk National University BK21 program (2024). This research was also supported by a grant of the MD-PhD/Medical Scientist Training Program through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea and by a grant from the Korea Research Institute of Bioscience and Biotechnology (KRIBB) Research Initiative Program (KGM5192531).

Data availability

The RNA-seq-generated datasets are available in the GEO public database under data series accession numbers GSE294703 and GSE294705 [16, 17]. In addition, the R functions for calculating the High-Frequency Recurrence Index (HfRI) and the trained machine learning prediction models are openly accessible at: <https://github.com/baeksw33/HighFrequencyRecurIndexNMIBC>. The clinical data and study protocol for this article can be obtained from the corresponding author.

Author's contributions

SJY had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. *Concept and design:* SKK, SJY. *Acquisition, analysis, or*

interpretation of data: SWB, YJB, XMP, OK, SM, KK, HWK, WTK SKK, SJY. *Drafting of the manuscript:* SWB, YJB, SKK, SJY. *Critical revision of the manuscript for important intellectual content:* All authors. *Statistical analysis:* SWB, YJB, SKK, JWS, IS SYK *Obtaining funding:* SWB, SKK, YJB, XMP, SJY. *Administrative, technical, or material support:* SYK, YHC, SKM, WJK, YDC. *Supervision:* SKK, SJY. All authors read and approved the final manuscript

Competing interests

The authors declare that they have no competing interests.

Ethics approval and consent to participate

All patient-derived tissues were collected after surgery at Chungbuk National University Hospital (CBNUH) in accordance with institutional procedures. The study was approved by the Institutional Review Board of Chungbuk National University and CBNUH (IRB approval numbers: CBNU-202309-BR-0201 and GR2010-12-010, respectively), and written informed consent was obtained from all participants prior to enrollment. Specimens were provided by CBNUH, a member of the National Biobank of Korea supported by the Ministry of Health, Welfare and Family Affairs, and all biobank samples were obtained under IRB-approved protocols with informed consent.

Figure legends

Figure 1. Clinical relevance and predictive performance of High-frequency Recurrence Index (HfRI) in UROMOL2021 cohort.

(A) Kaplan-Meier analysis of recurrence-free survival (RFS) with in the UROMOL2021 cohort (n=535), shows that patients with HfRI-high group (red) experienced significantly poorer prognosis compared to the HfRI-low group (blue) ($p<0.001$ by the log-rank test). (B) Kaplan-Meier analysis restricted to high-frequency recurrence (HfR; ≥ 2 recurrence) patients (n=267), excluding single recurrence cases, also revealed poorer prognosis in the HfRI-high group ($p<0.001$ by the log-rank test). (C) The bar plot showing recurrence frequency stratified by HfRI group. The HfRI-high group had a significantly higher proportion of patients with ≥ 2 recurrence compared to the HfRI-low group ($p=0.001$ by χ^2 test). (D) Area Under Receiver Operating Characteristic (AUROC) analysis for predicting high-frequency recurrence (≥ 2 recurrences) by HfRI alone (A model, AUC=0.736, c-index=0.673) and HfRI combined with EORTC (B model, AUC=0.746, c-index=0.695), and HfRI combined with both EORTC and T1High (C model, AUC=0.750, c-index=0.696).

Figure 2. Clinical relevance and predictive performance of HfRI in Total cohort.

(A) Kaplan-Meier analysis of recurrence-free survival (RFS) with in the total cohort (n=485), excluding patients with only a single recurrence, shows that patients with HfRI-high group (red) experienced significantly worse RFS compared to the HfRI-low group (blue) ($p<0.001$ by the log-rank test). (B) Kaplan-Meier analysis restricted to high-frequency recurrence (HfR; ≥ 2 recurrence) patients who received BCG therapy (n=148), excluding single recurrence cases, also revealed poorer prognosis in the HfRI-high group ($p=0.02$ by the log-rank test). (C) The bar plot showing recurrence frequency stratified by HfRI group. The HfRI-high group had a significantly higher proportion of patients with ≥ 2 recurrence compared to the HfRI-low group ($p=0.001$ by χ^2 test). (D) AUROC analysis for predicting high-frequency recurrence (≥ 2 recurrences) by HfRI (A model, AUC=0.694, c-index=0.648) and HfRI combined with EORTC (B model, AUC=0.695, c-index=0.652), and HfRI

combined with both EORTC and T1High (C model, AUC=0.708, c-index=0.661).

ARTICLE IN PRESS

References

1. Bray F, Laversanne M, Sung H, Ferlay J, Siegel RL, Soerjomataram I, Jemal A: **Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries**. *CA Cancer J Clin* 2024, **74**(3):229-263.
2. Tran L, Xiao JF, Agarwal N, Duex JE, Theodorescu D: **Advances in bladder cancer biology and therapy**. *Nat Rev Cancer* 2021, **21**(2):104-121.
3. Kamat AM, Hahn NM, Efstathiou JA, Lerner SP, Malmstrom PU, Choi W, Guo CC, Lotan Y, Kassouf W: **Bladder cancer**. *Lancet* 2016, **388**(10061):2796-2810.
4. Babjuk M, Burger M, Capoun O, Cohen D, Comperat EM, Dominguez Escrig JL, Gontero P, Liedberg F, Masson-Lecomte A, Mostafid AH *et al*: **European Association of Urology Guidelines on Non-muscle-invasive Bladder Cancer (Ta, T1, and Carcinoma in Situ)**. *Eur Urol* 2022, **81**(1):75-94.
5. Sharma V, Chamie K, Schoenberg M, Lee VS, Fero K, Lec P, Munneke JR, Aaronson DS, Kushi LH, Quesenberry CP *et al*: **Natural History of Multiple Recurrences in Intermediate-Risk Non-Muscle Invasive Bladder Cancer: Lessons From a Prospective Cohort**. *Urology* 2023, **173**:134-141.
6. Sylvester RJ, van der Meijden AP, Oosterlinck W, Witjes JA, Bouffieux C, Denis L, Newling DW, Kurth K: **Predicting recurrence and progression in individual patients with stage Ta T1 bladder cancer using EORTC risk tables: a combined analysis of 2596 patients from seven EORTC trials**. *Eur Urol* 2006, **49**(3):466-465; discussion 475-467.
7. Lindskrog SV, Prip F, Lamy P, Taber A, Groeneveld CS, Birkenkamp-Demtroder K, Jensen JB, Strandgaard T, Nordentoft I, Christensen E *et al*: **An integrated multi-omics analysis identifies prognostic molecular subtypes of non-muscle-invasive bladder cancer**. *Nat Commun* 2021, **12**(1):2301.
8. Hurst CD, Cheng G, Platt FM, Castro MAA, Marzouka NS, Eriksson P, Black EVI, Alder O, Lawson ARJ, Lindskrog SV *et al*: **Stage-stratified molecular profiling of non-muscle-invasive bladder cancer enhances biological, clinical, and therapeutic insight**. *Cell Rep Med* 2021, **2**(12):100472.
9. Baek SW, Mun JY, Jang IH, Yang GE, Jeong MS, Kim SK, Nam JK, Chu IS, Leem SH: **YAP1 activation is associated with the progression and response to immunotherapy of non-muscle invasive bladder cancer**. *EBioMedicine* 2022, **81**:104092.
10. Kim SK, Byun YJ, Park SH, Piao XM, Zheng CM, Moon S, Kim K, Song SJ, Kang HW, Kim WT *et al*: **A 23-Gene Prognostic Index Predicts Progression and Bacillus Calmette-Guerin Response in Non-muscle-invasive Bladder Cancer**. *Eur Urol* 2024, **85**(4):400-402.
11. Song BN, Kim SK, Mun JY, Choi YD, Leem SH, Chu IS: **Identification of an immunotherapy-responsive molecular subtype of bladder cancer**. *EBioMedicine* 2019, **50**:238-245.
12. Dyrskjot L, Zieger K, Real FX, Malats N, Carrato A, Hurst C, Kotwal S, Knowles M, Malmstrom PU, de la Torre M *et al*: **Gene expression signatures predict outcome in non-muscle-invasive bladder carcinoma: a multicenter validation study**. *Clin Cancer Res* 2007, **13**(12):3545-3551.
13. Lee JS, Leem SH, Lee SY, Kim SC, Park ES, Kim SB, Kim SK, Kim YJ, Kim WJ, Chu IS: **Expression signature of E2F1 and its associated genes predict superficial to invasive progression of bladder tumors**. *J Clin Oncol* 2010, **28**(16):2660-2667.
14. Kim WJ, Kim EJ, Kim SK, Kim YJ, Ha YS, Jeong P, Kim MJ, Yun SJ, Lee KM, Moon SK *et al*: **Predictive value of progression-related gene classifier in primary non-muscle invasive bladder cancer**. *Mol Cancer* 2010, **9**:3.
15. Kim SK, Roh YG, Park K, Kang TH, Kim WJ, Lee JS, Leem SH, Chu IS: **Expression signature defined by FOXM1-CCNB1 activation predicts disease recurrence in non-muscle-invasive bladder cancer**. *Clin Cancer Res* 2014, **20**(12):3233-3243.
16. Baek SW, Byun YJ: **High-frequency recurrence index for predicting multiple recurrences in non-muscle-invasive bladder cancer (Discovery cohort)**. Gene Expression Omnibus <https://identifiers.org/geo:GSE294703> (2026).
17. Baek SW, Byun YJ: **High-frequency recurrence index for predicting multiple recurrences in non-muscle-invasive bladder cancer (Validation cohort)**. Gene Expression Omnibus <https://identifiers.org/geo:GSE294705> (2026).
18. 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* 2013, **29**(1):15-21.
19. Lindskrog SV: **UROMOL 2020 - RNA-seq data**. European Genome-phenome Archive (EGA) <https://identifiers.org/ega.dataset:EGAD00001006656> (2020).
20. Dyrskjot L: **Gene expression signatures predict outcome in non-muscle invasive bladder carcinoma - a multi-center validation study**. Gene Expression Omnibus <https://identifiers.org/geo:GSE5479> (2006).
21. Leem S CI, Kim Y, Kim S, Kim W: **Predeictive Value of Prognosis-Related Gene Expression Study**

22. **in Primary Bladder Cancer.** Gene Expression Omnibus <https://identifiers.org/geo:GSE13507> (2008). Song B KS, Leem S, Chu I. **Comprehensive molecular classification of bladder cancer reveals distinct prognostic subgroups with different sensitivities to immunotherapy.** Gene Expression Omnibus <https://identifiers.org/geo:GSE120736> (2018).
23. CD H, MA K. **Expression analysis of human stage Ta and stage T1 bladder tumours.** Gene Expression Omnibus <https://identifiers.org/geo:GSE163209> (2020).
24. Hanzelmann S, Castelo R, Guinney J: **GSVA: gene set variation analysis for microarray and RNA-seq data.** *BMC Bioinformatics* 2013, **14**:7.
25. Soliman NA, Yussif SM: **Ki-67 as a prognostic marker according to breast cancer molecular subtype.** *Cancer Biol Med* 2016, **13**(4):496-504.
26. Zhang G, Tai P, Fang J, Chen A, Chen X, Cao K: **Molecular subtypes based on centrosome-related genes can predict prognosis and therapeutic responsiveness in patients with low-grade gliomas.** *Front Oncol* 2023, **13**:1157115.
27. Lee J, Lee YJ, Bae SJ, Baek SH, Kook Y, Cha YJ, Lee JW, Son BH, Ahn SH, Lee HJ *et al*: **Ki-67, 21-Gene Recurrence Score, Endocrine Resistance, and Survival in Patients With Breast Cancer.** *JAMA Netw Open* 2023, **6**(8):e2330961.
28. Sriramulu S, Thoidingjam S, Chen WM, Hassan O, Siddiqui F, Brown SL, Movsas B, Green MD, Davis AJ, Speers C *et al*: **BUB1 regulates non-homologous end joining pathway to mediate radioresistance in triple-negative breast cancer.** *bioRxiv* 2024.
29. Zheng X, Liu X, Lei Y, Wang G, Liu M: **Glypican-3: A Novel and Promising Target for the Treatment of Hepatocellular Carcinoma.** *Front Oncol* 2022, **12**:824208.
30. Kohli K, Pillarisetty VG, Kim TS: **Key chemokines direct migration of immune cells in solid tumors.** *Cancer Gene Ther* 2022, **29**(1):10-21.
31. Ayers M, Lunceford J, Nebozhyn M, Murphy E, Loboda A, Kaufman DR, Albright A, Cheng JD, Kang SP, Shankaran V *et al*: **IFN-gamma-related mRNA profile predicts clinical response to PD-1 blockade.** *J Clin Invest* 2017, **127**(8):2930-2940.
32. P. Gontero (Chair) EC, J.L. Dominguez Escrig,, F. Liedberg PM, A. Masson-Lecomte, A.H. Mostafid,, B.W.G. van Rhijn MR, T. Seisen,, S.F. Shariat ENX, Patient Advocates: I. Gürses Andersson RW, Guidelines Associates: O. Capoun BP, B.P. Rai FS, V. Soukup, Bezuidenhout GOC: **EAU Guidelines on Non-muscle-Invasive Bladder-Cancer.** In: *The EAU Annual Congress.* Milan; 2023.
33. Botteman MF, Pashos CL, Redaelli A, Laskin B, Hauser R: **The health economics of bladder cancer: a comprehensive review of the published literature.** *Pharmacoeconomics* 2003, **21**(18):1315-1330.
34. Mariathasan S, Turley SJ, Nickles D, Castiglioni A, Yuen K, Wang Y, Kadel EE, III, Koepfen H, Astarita JL, Cubas R *et al*: **TGFbeta attenuates tumour response to PD-L1 blockade by contributing to exclusion of T cells.** *Nature* 2018, **554**(7693):544-548.
35. Chakravarthy A, Khan L, Bensler NP, Bose P, De Carvalho DD: **TGF-beta-associated extracellular matrix genes link cancer-associated fibroblasts to immune evasion and immunotherapy failure.** *Nat Commun* 2018, **9**(1):4692.
36. Zheng A, Wei Y, Zhao Y, Zhang T, Ma X: **The role of cancer-associated mesothelial cells in the progression and therapy of ovarian cancer.** *Front Immunol* 2022, **13**:1013506.
37. Castiglioni A, Yang Y, Williams K, Gogineni A, Lane RS, Wang AW, Shyer JA, Zhang Z, Mittman S, Gutierrez A *et al*: **Combined PD-L1/TGFbeta blockade allows expansion and differentiation of stem cell-like CD8 T cells in immune excluded tumors.** *Nat Commun* 2023, **14**(1):4703.

Figure 1

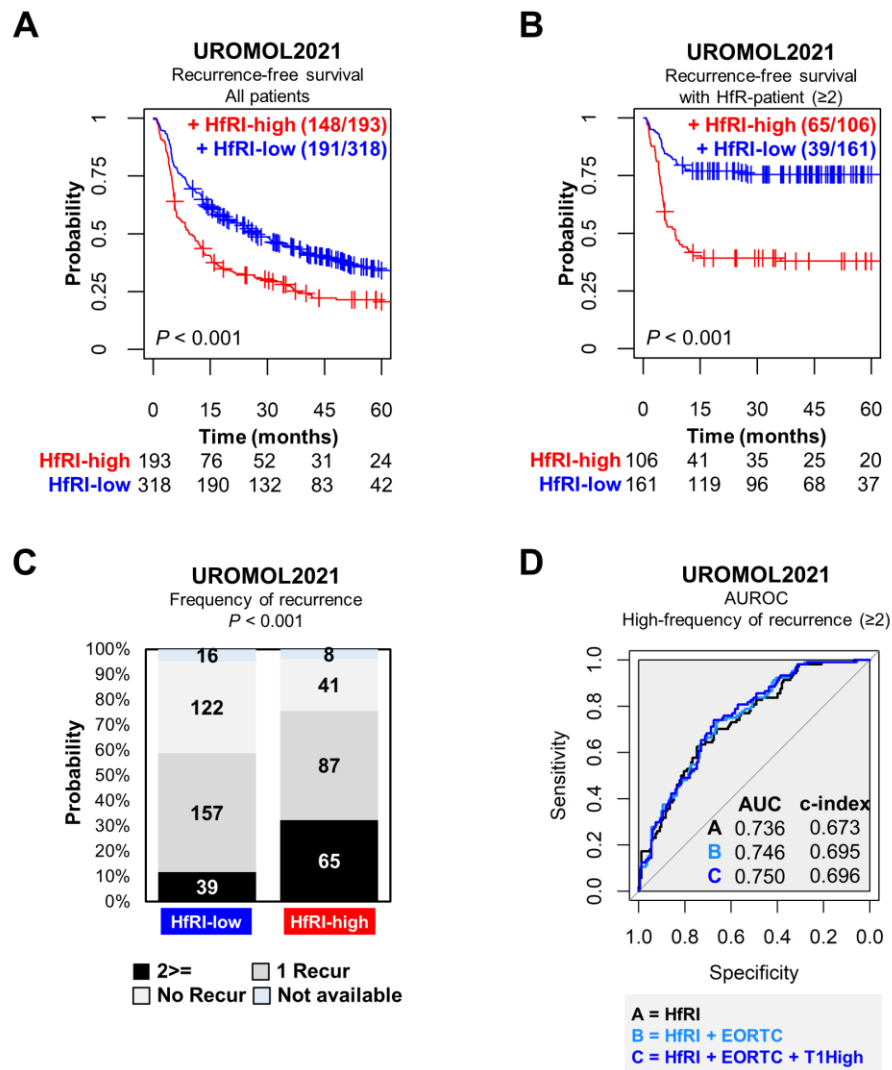


Figure 2

