

## Research Paper

# HES4 is a potential biomarker for bladder cancer: a Mendelian randomization study

Zhiyong Tan<sup>1,2,3#</sup>, Xiaorong Chen<sup>4#</sup>, Haihao Li<sup>1,2,3#</sup>, Yinglong Huang<sup>1,2,3#</sup>, Shi Fu<sup>1,2,3</sup>, Mingxia Ding<sup>1,2,3</sup>, Jiansong Wang<sup>1,2,3</sup>✉, and Haifeng Wang<sup>1,2,3</sup>✉

1. Department of Urology, The Second Affiliated Hospital of Kunming Medical University, No. 347, Dianmian Street, Wuhua District, Kunming, 650101, Yunnan, People's Republic of China.
2. Urological disease clinical medical center of Yunnan province, The Second Affiliated Hospital of Kunming Medical University, No. 347, Dianmian Street, Wuhua District, Kunming, 650101, Yunnan, People's Republic of China.
3. Scientific and Technological Innovation Team of Basic and Clinical Research of Bladder Cancer in Yunnan Universities, The Second Affiliated Hospital of Kunming Medical University, No. 347, Dianmian Street, Wuhua District, Kunming, 650101, Yunnan, People's Republic of China.
4. Department of Kidney Transplantation, The Third Hospital of Sun Yat-Sen University, Guangzhou, People's Republic of China.

#These authors contributed equally to this work.

✉ Corresponding authors: Jiansong Wang, Department of Urology, The Second Affiliated Hospital of Kunming Medical University, Yunnan Institute of Urology, Kunming 650101, China. E-mail: wangjiansong@kmmu.edu.cn and Haifeng Wang, Department of Urology, The Second Affiliated Hospital of Kunming Medical University, Yunnan Institute of Urology, Kunming 650101, China. E-mail: t13085348360@163.com.

© The author(s). This is an open access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>). See <http://ivyspring.com/terms> for full terms and conditions.

Received: 2023.11.27; Accepted: 2024.01.11; Published: 2024.01.21

## Abstract

**Background:** Patients with bladder cancer (BLCA) have a poor prognosis and little progress has been made in treatment. Therefore, the purpose of this work was to employ Mendelian randomization (MR) and transcriptome analysis to identify a novel biomarker that could be used to reliably diagnose BLCA.

**Methods:** TCGA-BLCA and GSE121711 datasets were obtained from public databases. Genome-wide association study (GWAS) data of BLCA outcome (373,295 samples containing 9,904,926 single nucleotide polymorphisms) were obtained through the IEU OpenGWAS database. Differentially expressed genes were applied as exposure factors, and MR analysis was performed to identify genes that had a causal relationship with BLCA. Then, the patients were divided into high and low expression groups according to the expression levels of candidate genes, and genes with survival differences were identified. Univariate and multivariate Cox regression were used to investigate the prognostic value of the expression of these genes. A nomogram was constructed based on independent prognostic factors, and we analyzed the functions and pathways associated with the identified genes as well as their relationship with the immune microenvironment.

**Results:** HES4 was identified as a biomarker. HES4 status, age, and stage were identified as independent prognostic factors, and an excellent nomogram was established. Bioinformatic analysis suggested that HES4 might be associated with the activation of the immune response, bone development, and cancer pathways. The BLCA samples were divided into high and low HES4 groups. The stromal score and 33 immune cells were remarkably different between the two groups, with HES4 expression being negatively correlated with macrophages and mast cells, and positively correlated with eosinophils and central memory CD4+ T cells. Finally, HES4 was up-regulated in cancer samples in both TCGA-BLCA and GSE121711 datasets.

**Conclusion:** This study identified HES4 as an independent prognostic factor for BLCA outcome based on MR and transcriptome analysis, which provides useful information for future research on and treatment of BLCA.

Keywords: Bladder cancer; Mendelian randomization; Biomarker; Tumor immune microenvironment

## 1. Introduction

Bladder cancer (BLCA) is a common malignant tumor of the urological system. It occurs more frequently in men for whom it is the sixth most prevalent cancer [1]. Currently identified risk factors for BLCA are smoking and occupational exposures.

Recent studies have revealed that specific dietary factors, microbiota imbalances, gene-external risk factor interactions, exposure to diesel exhaust emissions, and pelvic radiotherapy may also be contributing factors [2]. Depending on the depth of

tumor infiltration, it can be classified as non-muscle invasive bladder cancer (NMIBC) and muscle-invasive bladder cancer (MIBC). Approximately 75% of patients have NMIBC at the time of the initial visit [3], and the 5-year overall survival (OS) rate is nearly 90% [4]. Approximately 10–30% of NMIBC patients progress to MIBC with their 5-year OS decreasing to 60–70% [4,5], and 10% of these patients develop metastasis beyond the bladder, decreasing the 5-year OS to 5–30%. Apart from traditional therapies involving surgery, chemotherapy, and radiotherapy, immunotherapy has started to be used for the treatment of patients with advanced BLCA. However, only a limited number of patients benefit from immunotherapy [6], with one of the crucial reasons for this being the lack of biomarkers available to guide diagnosis and treatment. Although Chen et al. identified urinary tumor DNA methylation assessment as a novel method for early diagnosis of BLCA [7], it is still urgent to study the underlying molecular mechanisms of BLCA occurrence and progression in order to identify more sensitive and precise biomarkers. This will facilitate the early diagnosis and treatment of BLCA, which will help to increase the survival rate and improve the prognosis of patients.

Mendelian randomization (MR) uses genetic variants as instrumental variables (IVs), and the random assignment of effect alleles is similar to the randomization process in randomized controlled trials, which strengthens the inference of causality between exposure and outcome [8–10]. The strength of MR is that the method minimizes the interference of confounding factors such as environment and adaptation, and also reduces reverse causality since the genetic variation used to substitute for the effect of exposure cannot be changed in turn by the occurrence and progression of the outcome [11,12]. The association of coffee intake, alcohol consumption, smoking, human papillomavirus (HPV) infection, and benign prostatic hyperplasia (BPH) with the development of BLCA has been investigated in previous MR studies, with results suggesting that smoking, HPV infection, and BPH may be causally associated with BLCA, whereas coffee intake and alcohol consumption are not [13–16]. The European Association of Urology stated that there is currently insufficient evidence to support the use of tumor mutation burden (TMB), gene expression signature or molecular typing in the management of patients with BLCA [17]. Transcriptomic MR studies may help to change this by identifying genes causally associated with BLCA.

In this study, MR analysis of the BLCA transcriptome was performed to identify novel causal

BLCA genes. The molecular mechanisms by which these genes might impact BLCA development and progression were investigated by enrichment analysis, ceRNA analysis, immune infiltration analysis, and targeted drug prediction.

## 2. Materials and methods

### 2.1 Data sources

The Cancer Genome Atlas (TCGA) bladder cancer dataset (TCGA-BLCA) was obtained from the TCGA database (<https://www.cancer.gov/tcga>), which contained 397 BLCA samples with complete survival information. GSE121711 (GPL17586) was obtained from the Gene Expression Omnibus (GEO) database (<https://www.ncbi.nlm.nih.gov/geo>), which contained 8 BLCA samples and 10 normal samples. Genome-wide association study (GWAS) data of BLCA outcomes (ieub-4874; 373,295 samples containing 9,904,926 single nucleotide polymorphisms (SNPs)) and independent SNPs were obtained through the IEU OpenGWAS database (<https://gwas.mrcieu.ac.uk/>).

### 2.2 Differentially expressed analysis

Differentially-expressed genes (DEGs) were obtained between tumor and normal samples using DESeq2 (v3.44.3) in the R package [18]. An adjusted  $P < 0.05$  and  $|\log_2FC| > 1$  were used as cutoffs to define differential expression. Univariate Cox regression was used to identify DEGs associated with outcome, with  $P < 0.05$  considered statistically significant. Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) was used to explore the functions and associated pathways of the identified genes using clusterProfiler (v3.16.0) in R [19].

### 2.3 Selecting IVs for MR analysis

To ensure valid MR analysis, three basic assumptions should be satisfied: (1) IVs are closely related to exposure factors, (2) IVs cannot be associated with any confounding factors, and (3) exposure is the only way that genetic variation can affect the outcome. For univariate MR analysis, SNPs were used as IVs. The R package TwoSampleMR (v0.5.6) [20] was used to read the exposure factors and screen the IVs. SNPs were selected using the genome-wide significance threshold  $P < 5 \times 10^{-8}$ . Appropriate SNPs were then retained using a linkage imbalance threshold of  $r^2 = 0.001$  in the Euro 1000 genome reference panel. Finally, SNPs that were associated with the results but with  $P < 5 \times 10^{-8}$  were excluded [21]. Palindromic SNPs with intermediate allele frequencies were eliminated when combining exposure and outcome datasets.

## 2.4 MR analysis

MR analysis was used to explore the causal effect between exposure factors (DEGs identified by the Cox analysis) and outcome (BLCA). In this study, the MR-Egger, weighted median, inverse variance weighted (IVW), simple mode, and weighted mode [22] methods were used for univariate MR analysis. The IVW method was used as the main MR analysis to identify genes that had a causal relationship with BLCA. A P value cutoff of 0.05, was used to identify causal relationships. A scatter plot was used to determine the correlation between exposure factors and outcome, a forest plot was used to determine the predictive exposure factors of each SNP site for the diagnosis of the outcome, and a funnel plot was used to evaluate whether MR was in line with Mendel's second law of randomization.

## 2.5 Sensitivity analysis

Sensitivity analysis was used to evaluate the reliability of MR analysis. Firstly, the Q value of the heterogeneity test was more than 0.05, indicating that there was no heterogeneity. Secondly, the P-value of the horizontal pleiotropy test was greater than 0.05, indicating that there were no confounding factors. Finally, the leave-one-out (LOO) test was used to determine whether the remaining SNPs had an impact on the overall analysis, even if one SNP was removed.

## 2.6 Identification of key genes

Univariate Cox regression and multivariate Cox regression were used to analyze the genes that had a causal relationship with BLCA, and genes whose Mendel risk values and univariate Cox risk values were in the same direction were selected as candidate genes for subsequent analysis. The expression levels of these genes and survival information of cancer patients were extracted from the TCGA dataset for analysis of overall survival (OS), and Kaplan-Meier (K-M) survival curves were constructed. Patients were divided into high and low expression groups based on the expression level of the candidate gene. Candidate genes with significant differences in survival were selected as biomarkers for subsequent analysis.

## 2.7 Independent prognostic analysis

Clinicopathological factors and biomarker expression levels in TCGA-BLCA samples were used to explore the independent prognostic value of the biomarkers by univariate Cox analysis. A nomogram was constructed based on significant factors, and patient survival at 1, 3, and 5 years was predicted based on the total score, with the higher the score, the lower the survival. Calibration curve and decision

curve analysis (DCA) were used to evaluate the predictive ability of the nomogram.

## 2.8 Single gene GSEA and co-expression gene enrichment analysis

To explore the regulatory pathways and biological functions related to biomarkers expression, a single gene GSEA analysis was performed. Firstly, the correlation coefficients between the expression levels of all genes and biomarkers were calculated as the sequencing criteria, and GSEA enrichment analysis was performed using clusterProfiler (v3.16.0) [23] in the R package. The screening criteria were  $|NES| > 1$ , adj.  $P < 0.05$ . To further understand the functions and pathways associated with genes that were co-expressed with the biomarkers, LinkedOmics was used to identify co-expressed genes, their correlation was assessed using Spearman analysis, and GO and KEGG analyses were conducted. The screening criterion was adj.  $P < 0.05$ .

## 2.9 Estimation of tumor immune micro-environment (TIME)

A series of correlation analyses were used to assess the differences in the TIME between high and low expression groups. The R package estimate (v1.0.13) [24] was used to calculate the immune score, matrix score, ESTIMATE score, and tumor purity of each sample, and the association between them was examined by calculating the Spearman correlation coefficient. Seven different approaches were used to assess immune cell differences between high and low expression groups (TIMER, CIBERSORT, CIBERSORT-ABS, quantTIseq, MCP-counter, xCELL, and EPIC). TIDE is a computational method for predicting immune checkpoint blocking (ICB) responses [25]. From the RNA expression profile, TIDE prediction scores were calculated to predict the likelihood that BLCA patients would respond to immunotherapy. The lower the TIDE score, the lower the likelihood of immune escape. In addition, subMap analysis was performed to predict differences in response to anti-PD-1 or anti-CTLA-4 immunotherapy between high and low expression groups. Finally, as immune checkpoint inhibitors (ICI) are used clinically for cancer immunotherapy, Spearman correlation analysis was used to investigate the relationship between the identified biomarkers and expression of genes associated with immune checkpoints.

## 2.10 Prediction of chemotherapy drug sensitivity and molecular docking

To attempt to identify high-potency anti-BLCA drugs, the R pRRopheticPredict package (v5.0) was used to calculate the half-inhibitory concentrations



(IC<sub>50</sub>) of 138 drugs in the CGP database. The difference in the IC<sub>50</sub> values between the high and low expression groups was then calculated. Spearman correlation analysis was used to assess the correlation between the IC<sub>50</sub> values and the expression of the identified biomarkers. To identify a potential novel therapeutic target for an identified drug, it was docked to the structure of HES4. The protein structure was obtained from the PDB [26], and all small molecules (including water) were removed. Protein hydrogenation and charge calculations were performed using AutoDock. Drug structures were retrieved from the PubChem database (<https://pubchem.ncbi.nlm.nih.gov>) [27] and charge-balanced, rotatable bond checks were performed using AutoDock. A range of docking boxes were selected based on the receptor active site. AutoDock vina was used to perform protein-ligand docking, and the structure with the lowest binding free energy was selected. PyMol software was used for visualization.

### 2.11 Evaluation of genetic mutations, tumor mutation burden (TMB), and copy number variation (CNV)

To identify the gene mutations in the high and low expression group, somatic mutation data were downloaded from the Genomic Data Commons (GDC) database [28], and mutations in the high and low expression groups were identified using maftools (v2.4.10) [29] in R. TMB is a quantitative biomarker that reflects the total number of mutations carried by tumor cells. To investigate differences in TMB between the high and low expression groups, we calculated TMB values and made boxplots. CNV in the TCGA-BLCA data were displayed on chromosome plots, and Genomic Identification of Significant Targets in Cancer (GISTIC) was used to analyze copy number loss or gain in all samples followed by a Wilcoxon test between the high and low expression groups to identify statistically significant differences.

### 2.12 Construction of a competing endogenous RNA (ceRNA) network

The miRWalk tool (<http://mirwalk.umm.uni-heidelberg.de/>) was used to predict biomarker-related micro RNAs (miRNAs). miRNA-related long non-coding RNAs (lncRNAs) were identified in Starbase (<http://starbase.sysu.edu.cn/>) using a threshold of clipExpNum > 1, and the ceRNA network was constructed.

### 2.13 Expression verification of biomarkers

The expression levels of key genes were verified

in the TCGA-BLCA and GSE121711 datasets, and results were displayed using boxplots.

### 2.14 Statistical analysis

Statistical analysis was performed using R software (v4.1.0). Differences between groups were analyzed using Wilcoxon tests and t-tests. \**P*-value < 0.05; \*\**P*-value < 0.005; \*\*\**P*-value < 0.0005; \*\*\*\**P*-value < 0.00005, represented a significant difference. In the GSEA, the thresholds used were |NES| > 1 and adj. *P* < 0.05.

## 3. Results

### 3.1 Identification of DEGs with differences in risk ratio

A total of 8467 DEGs were identified between BLCA and normal samples, with 3374 down-regulated and 5093 up-regulated (**Figure 1A-B**). Univariate Cox proportional hazard regression analysis was performed on all 8467 DEGs, with a total of 3678 giving significant differences in survival (*P* < 0.05) (**Supplementary Table 1**).

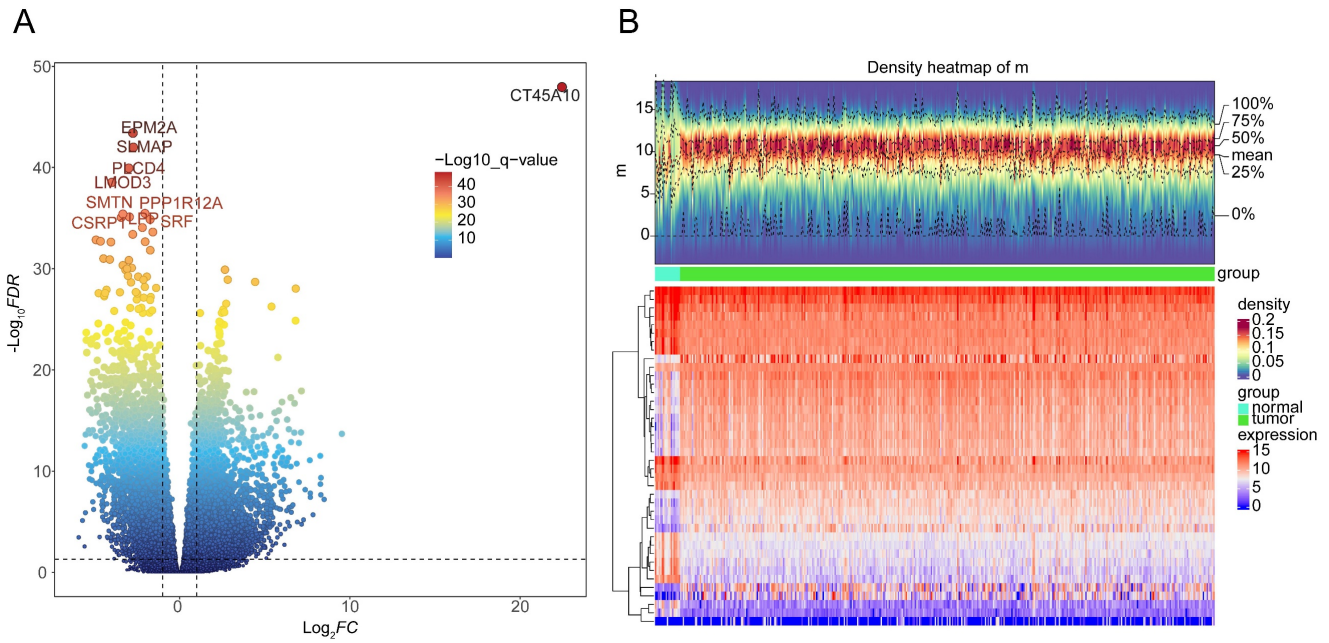
### 3.2 Functional enrichment and pathway analysis

A total of 9 KEGG pathways and 58 GO function entries were enriched in the set of 3678 genes identified above. The most significant KEGG pathways were herpes simplex virus infection, spliceosome, Fanconi anemia pathway, mRNA surveillance pathway, hypertrophic cardiomyopathy and RNA degradation (**Figure 2A**). The most significant GO terms were RNA splicing via spliceosome and tRNA processing (**Figure 2B**).

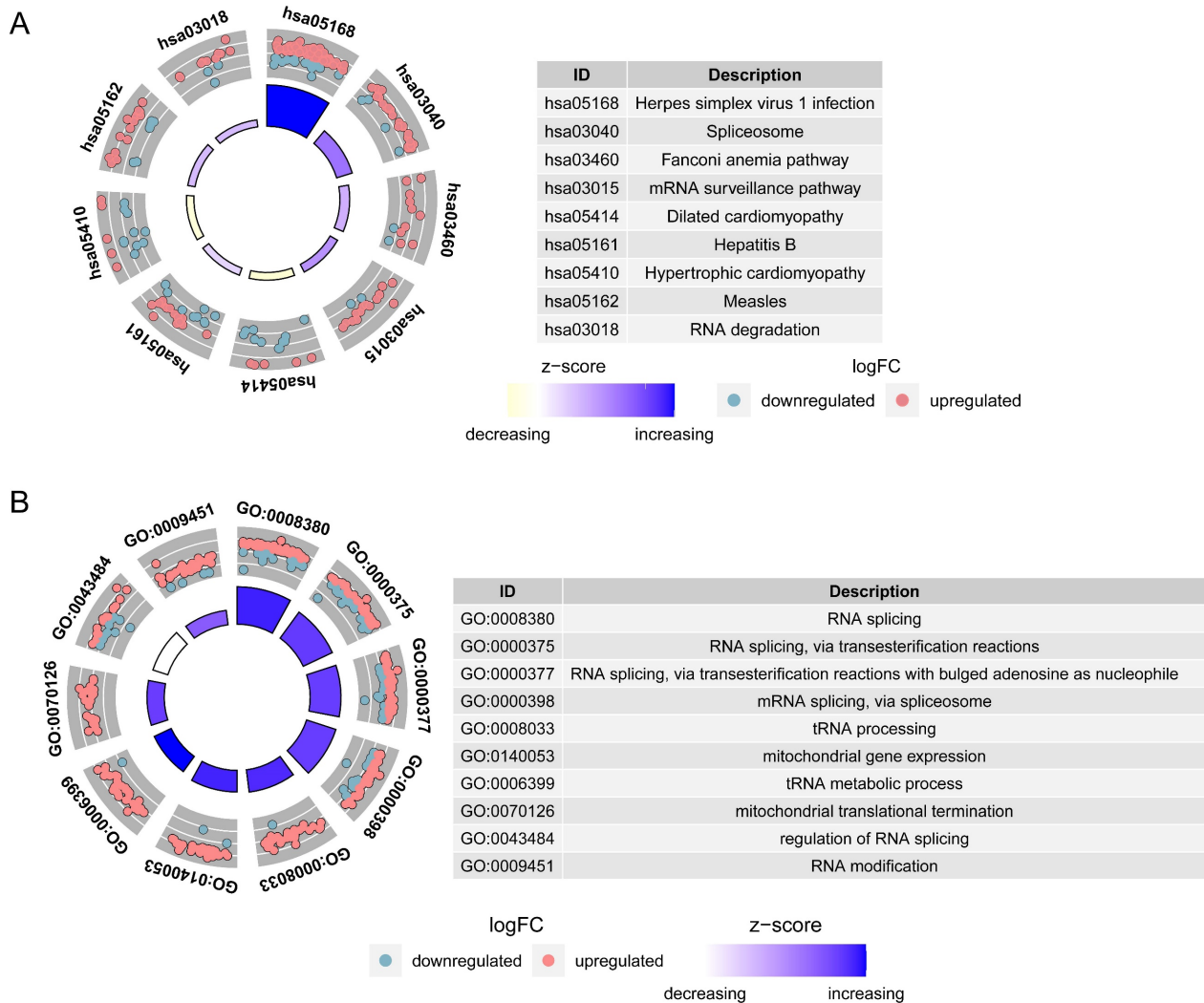
### 3.3 Identification of genes with a causal relationship with BLCA

The IVW results showed that 8 genes were causally associated with BLCA. These were *NSUN5P2* (*P* = 0.025), *GHRLOS* (*P* = 0.000), *TPM3P9* (*P* = 0.021), *HES4* (*P* = 0.021), *BATF2* (*P* = 0.002), *PLXNA4* (*P* = 0.019), *TGFBI* (*P* = 0.014) and *EPHB1* (*P* = 0.040). *HES4* (Odds ratio [OR] = 0.888), *PLXNA4* (OR = 0.919), and *EPHB1* (OR = 0.811) were safety factors for BLCA, and *NSUN5P2* (OR = 1.030), *GHRLOS* (OR = 1.011), *TPM3P9* (OR = 1.018), *BATF2* (OR = 1.046), and *TGFBI* (OR = 1.074) were risk factors (**Supplementary Table 2**). For *HES4*, the scatter plot shows that the IVW slope is negative with no intercept (**Figure 3A**). The forest map also demonstrates that *HES4* is a safety factor (**Figure 3B**). SNPs were randomly distributed on both sides of the IVW line, indicating that MR conformed to Mendel's second law (**Figure 3C**). The results for all other genes are shown in **Supplementary Figure 1-3**.

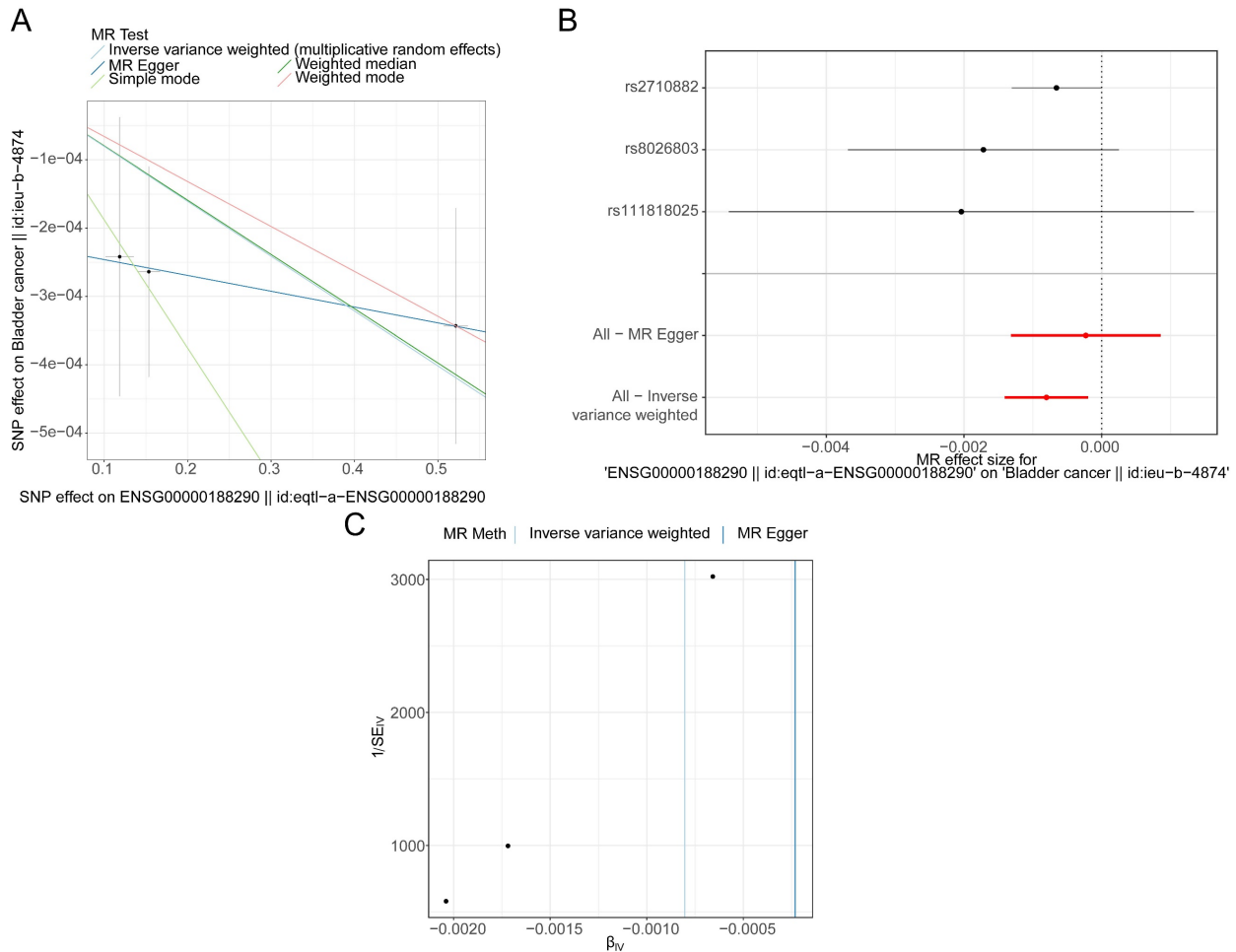




**Figure 1.** Identification of differentially expressed genes (DEGs) between BLCA and normal groups in the TCGA-BLCA dataset. **A** Volcano plot and **B** heatmap (density) plot illustrating the expression of 8467 DEGs.



**Figure 2.** Functional enrichment circle plots of 3678 DEGs with significant differences in survival in univariate Cox analysis. **A** Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis. **B** Gene Ontology (GO) enrichment analysis.



**Figure 3.** *HES4* has a causal relationship with BLCA. **A** Scatter plots for the relationship of *HES4* expression and BLCA. **B** Forest plot for the efficacy of *HES4* in predicting BLCA risk. **C** Funnel plot demonstrating a symmetrical distribution of single nucleotide polymorphisms of *HES4* in both inverse variance weighted and MR-Egger algorithms.

### 3.4 Identification of *HES4* as biomarker

Sensitivity analysis was used to assess the reliability of the MR results. The heterogeneity test demonstrated that there was no heterogeneity in the analysis ( $P < 0.05$ ) (Table 1), the horizontal pleiotropy test showed that there were no confounding factors in the study ( $P > 0.05$ ) (Table 2), and LOO revealed that there were no outlier SNPs, all of which suggest that the MR results were reliable (Figure 4A, Supplementary Figure 4). The forest maps of these 8 genes showed that the Mendelian risk values of *HES4* (MR:  $P = 0.009$ , OR = 0.9992; univariate Cox analysis: OR = 0.868) and *TGFBI* (MR:  $P = 0.016$ , OR = 1.0005; univariate Cox analysis: OR = 1.0867) were in the same direction as the univariate Cox risk values (Figure 4B). Finally, OS analysis was performed using the expression levels of *HES4* and *TGFBI* and the patient survival information (Figure 4C-D). *HES4*, with the more significant impact on survival prognosis, was selected to proceed with further analyses.

### 3.5 *HES4*, age, and stage are independent prognostic factors

The clinicopathological factors and *HES4* gene expression of TCGA-BLCA samples were used for Cox analysis. The results indicated that the N ( $P = 4.13 \times 10^{-8}$ ), stage ( $P = 5.21 \times 10^{-8}$ ), age ( $P = 3.68 \times 10^{-5}$ ) and *HES4* expression ( $P = 0.0005$ ) were significant prognostic factors (Table 3, Figure 5A). *HES4* expression, age, and stage had significant  $P$ -values in multivariate Cox analysis, indicating that they were independent prognostic factors (Table 4, Figure 5B). A nomogram was constructed based on *HES4* expression, clinical factors, and the 1-, 3-, and 5-year survival rates (Figure 5C). The calibration curve slope (c-index = 0.677) was close to 1, showing that the nomogram prediction was more accurate. The slopes for 1, 3, and 5 years were 0.6935, 0.2826, and 0.2122, respectively, indicating that the prediction was best at 1 year (Figure 5D). DCA showed a nomogram with the sole advantage of *HES4*, indicating that the nomogram has good forecasting ability (Figure 5E).

**Table 1.** Sensitivity analysis based on the heterogeneity test.

id.exposure	id.outcome	outcome	exposure	method	Q	Q_df	Q_pval
eqtl-a-ENSG00000106133	ieu-b-4874	Bladder cancer     id:ieu-b-4874	ENSG00000106133     id:eqtl-a-ENSG00000106133	MR Egger	8.78E-01	1	0.34878779
eqtl-a-ENSG00000106133	ieu-b-4874	Bladder cancer     id:ieu-b-4874	ENSG00000106133     id:eqtl-a-ENSG00000106133	Inverse variance weighted	9.59E-01	2	0.61917592
eqtl-a-ENSG00000120708	ieu-b-4874	Bladder cancer     id:ieu-b-4874	ENSG00000120708     id:eqtl-a-ENSG00000120708	MR Egger	10.0967612	15	0.81360960
eqtl-a-ENSG00000120708	ieu-b-4874	Bladder cancer     id:ieu-b-4874	ENSG00000120708     id:eqtl-a-ENSG00000120708	Inverse variance weighted	1.09E+01	16	0.81570841
eqtl-a-ENSG00000154928	ieu-b-4874	Bladder cancer     id:ieu-b-4874	ENSG00000154928     id:eqtl-a-ENSG00000154928	MR Egger	5.55E+00	8	0.69766993
eqtl-a-ENSG00000154928	ieu-b-4874	Bladder cancer     id:ieu-b-4874	ENSG00000154928     id:eqtl-a-ENSG00000154928	Inverse variance weighted	5.61E+00	9	0.77822306
eqtl-a-ENSG00000168062	ieu-b-4874	Bladder cancer     id:ieu-b-4874	ENSG00000168062     id:eqtl-a-ENSG00000168062	MR Egger	1.14701925	1	0.28417399
eqtl-a-ENSG00000168062	ieu-b-4874	Bladder cancer     id:ieu-b-4874	ENSG00000168062     id:eqtl-a-ENSG00000168062	Inverse variance weighted	1.46E+00	2	0.48239952
eqtl-a-ENSG00000188290	ieu-b-4874	Bladder cancer     id:ieu-b-4874	ENSG00000188290     id:eqtl-a-ENSG00000188290	MR Egger	2.96E-03	1	0.95659970
eqtl-a-ENSG00000188290	ieu-b-4874	Bladder cancer     id:ieu-b-4874	ENSG00000188290     id:eqtl-a-ENSG00000188290	Inverse variance weighted	1.54E+00	2	0.46330548
eqtl-a-ENSG00000221866	ieu-b-4874	Bladder cancer     id:ieu-b-4874	ENSG00000221866     id:eqtl-a-ENSG00000221866	MR Egger	1.86836791	1	0.17166212
eqtl-a-ENSG00000221866	ieu-b-4874	Bladder cancer     id:ieu-b-4874	ENSG00000221866     id:eqtl-a-ENSG00000221866	Inverse variance weighted	1.92E+00	2	0.38348767
eqtl-a-ENSG00000240288	ieu-b-4874	Bladder cancer     id:ieu-b-4874	ENSG00000240288     id:eqtl-a-ENSG00000240288	MR Egger	1.18E+00	2	0.55481913
eqtl-a-ENSG00000240288	ieu-b-4874	Bladder cancer     id:ieu-b-4874	ENSG00000240288     id:eqtl-a-ENSG00000240288	Inverse variance weighted	2.45E+00	3	0.48470037
eqtl-a-ENSG00000241015	ieu-b-4874	Bladder cancer     id:ieu-b-4874	ENSG00000241015     id:eqtl-a-ENSG00000241015	MR Egger	0.14033583	2	0.93223726
eqtl-a-ENSG00000241015	ieu-b-4874	Bladder cancer     id:ieu-b-4874	ENSG00000241015     id:eqtl-a-ENSG00000241015	Inverse variance weighted	1.58E+00	3	0.66433218

**Table 2.** Sensitivity analysis based on the horizontal pleiotropy test.

id.exposure	id.outcome	outcome	exposure	egger_intercept	se	pval
eqtl-a-ENSG00000106133	ieu-b-4874	Bladder cancer     id:ieu-b-4874	ENSG00000106133     id:eqtl-a-ENSG00000106133	-7.10E-05	0.000249604	0.82361746
eqtl-a-ENSG00000120708	ieu-b-4874	Bladder cancer     id:ieu-b-4874	ENSG00000120708     id:eqtl-a-ENSG00000120708	5.08E-05	5.67E-05	0.38472654
eqtl-a-ENSG00000154928	ieu-b-4874	Bladder cancer     id:ieu-b-4874	ENSG00000154928     id:eqtl-a-ENSG00000154928	-1.93E-05	7.79E-05	0.81020401
eqtl-a-ENSG00000168062	ieu-b-4874	Bladder cancer     id:ieu-b-4874	ENSG00000168062     id:eqtl-a-ENSG00000168062	-0.00061497	0.001181128	0.69439614
eqtl-a-ENSG00000188290	ieu-b-4874	Bladder cancer     id:ieu-b-4874	ENSG00000188290     id:eqtl-a-ENSG00000188290	-0.00022855	0.000179828	0.43223463
eqtl-a-ENSG00000221866	ieu-b-4874	Bladder cancer     id:ieu-b-4874	ENSG00000221866     id:eqtl-a-ENSG00000221866	-0.000167165	0.001037243	0.89827553
eqtl-a-ENSG00000240288	ieu-b-4874	Bladder cancer     id:ieu-b-4874	ENSG00000240288     id:eqtl-a-ENSG00000240288	-0.00021807	0.000193497	0.3767828
eqtl-a-ENSG00000241015	ieu-b-4874	Bladder cancer     id:ieu-b-4874	ENSG00000241015     id:eqtl-a-ENSG00000241015	-0.000197497	0.0001647	0.35327642

**Table 3.** Results for independent prognostic univariate COX analysis.

variable	coef	HR	HR.95L	HR.95H	pvalue
N	0.462116395	1.587430062	1.345801715	1.872440919	4.13E-08
STAGE	0.610093568	1.840603612	1.477634912	2.292732548	5.21E-08
age	0.035464194	1.036100549	1.018794898	1.05370016	3.68E-05
HES4	-0.187302277	0.829193053	0.745891558	0.921797695	0.000525474
gender	-0.122877782	0.884371742	0.622876769	1.25564705	0.492041845

### 3.6 Signaling mechanisms related to HES4

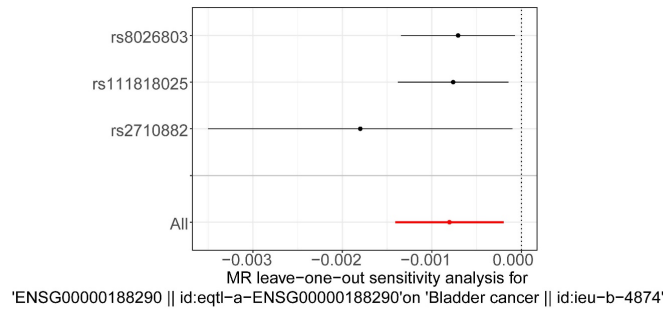
GSEA enrichment analysis was used to understand the functions and pathways associated with HES4. A total of 3822 GO entries contained HES4, including actomyosin structural organization, aerobic respiration, ATP synthesis coupled electron transport, and bone development (Figure 6A). 81 KEGG pathways contained HES4, including oxidative phosphorylation, Parkinson's disease, and cancer pathway (Figure 6B). Using LinkedOmics and Spearman correlation analysis, the 50 genes whose

expression were most strongly positively and negatively correlated with HES4 were found. Enrichment analysis of these 100 genes identified 4 KEGG pathways and 78 GO entries. The KEGG pathways were cell senescence, lysine degradation, sulfur relay system, and lipid and atherosclerosis. Cellular senescence and lipid and atherosclerosis pathways contained the most genes (Figure 6C). Enriched GO terms included peptidyl-serine modification, proteasome-mediated ubiquitin-dependent protein catabolism, peptidyl-threonine modification,

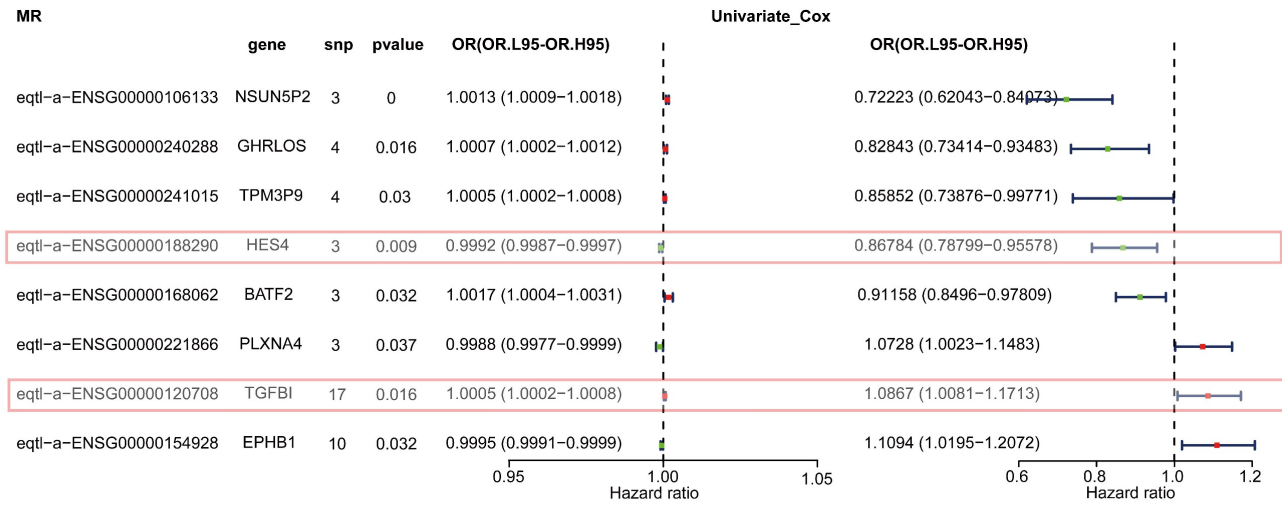


megakaryocyte development, histone methylation, and glycoprotein methylation. Peptidyl-serine modification contained the most genes (Figure 6D).

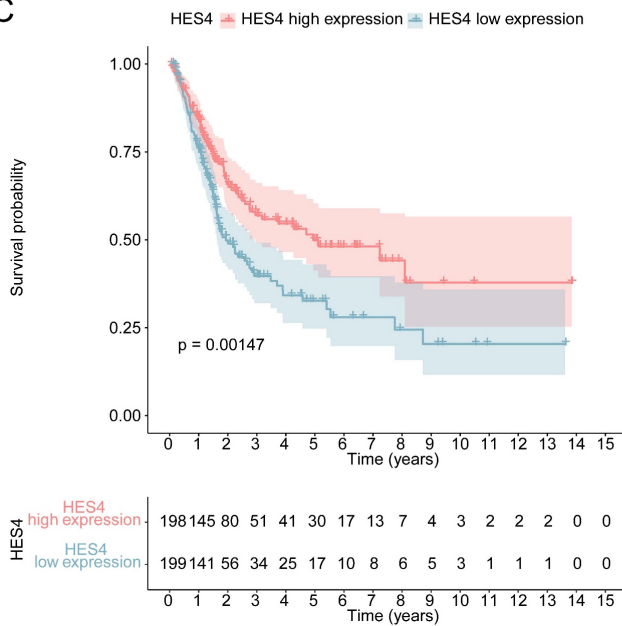
A



B MR



C



D

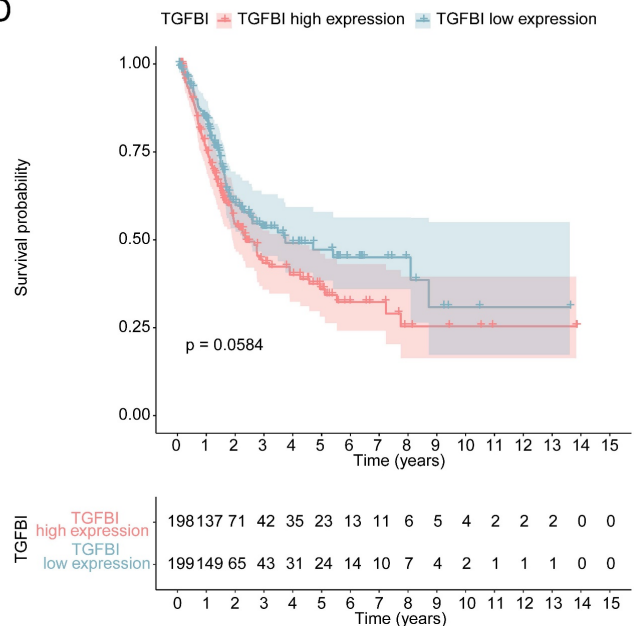
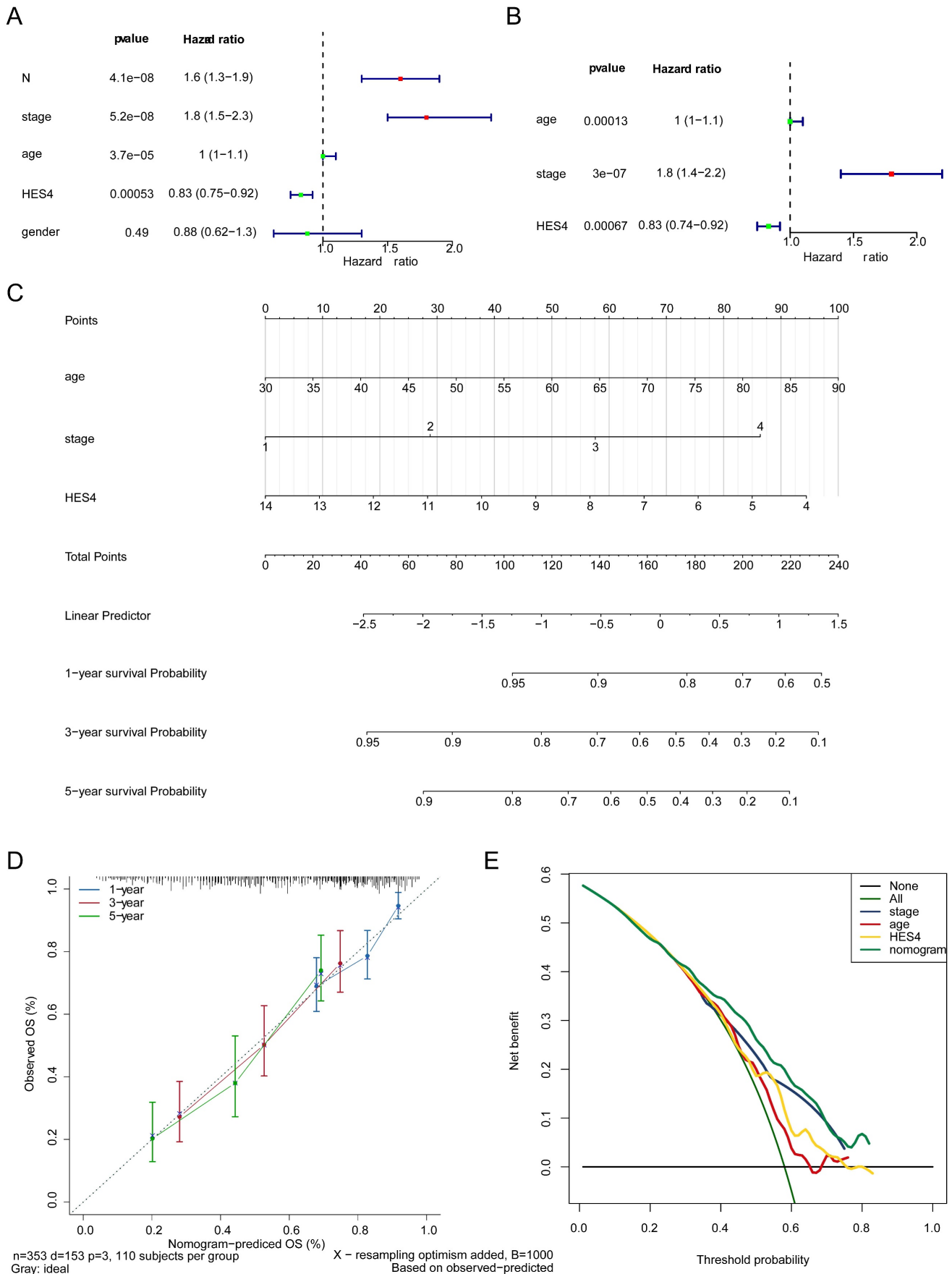
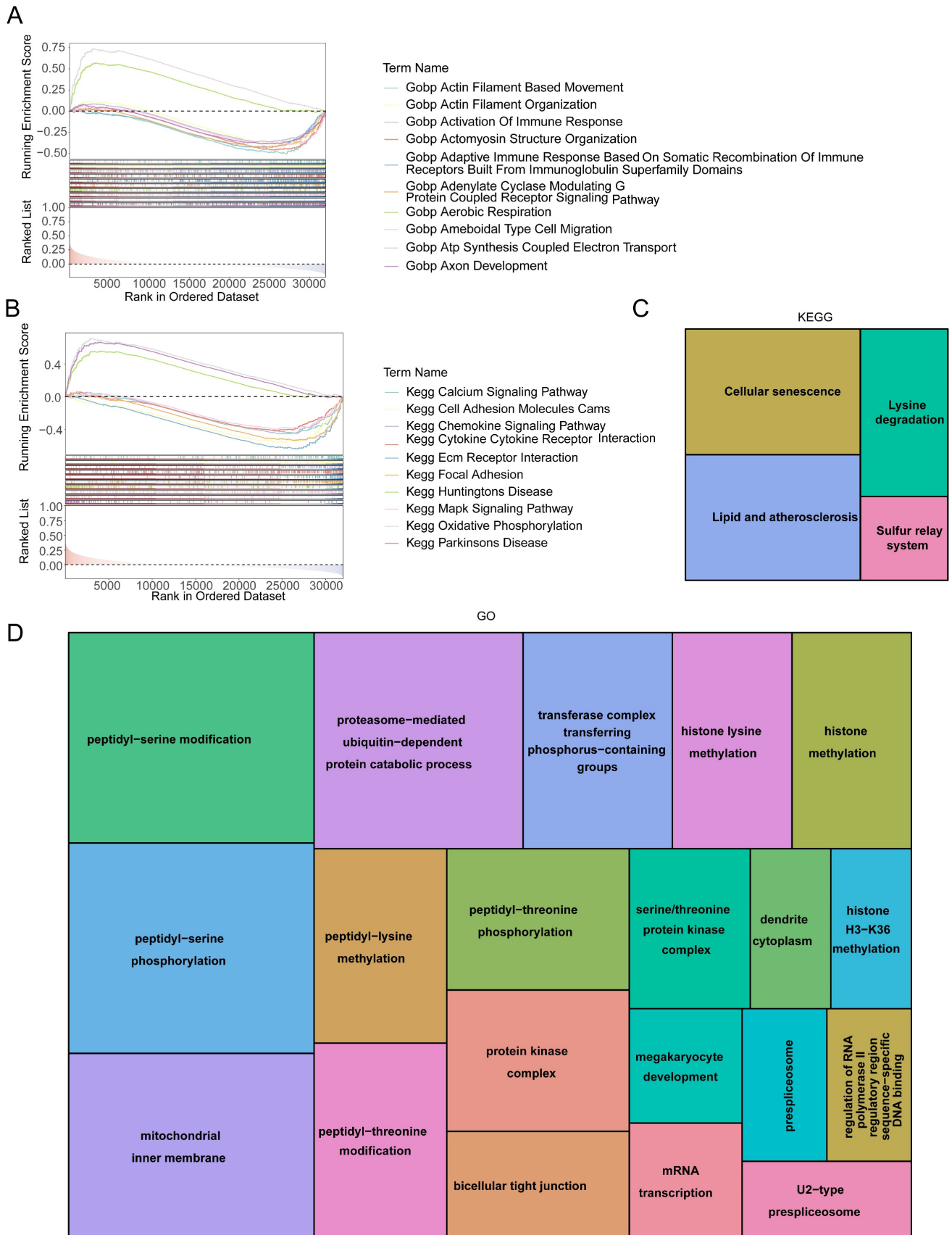


Figure 4. Sensitivity analysis of MR. A Forest plot of leave-one-out analysis of the relationship of HES4 and BLCA. B Forest plot for the univariate Cox analysis of eight genes with a causal relationship with BLCA. C Kaplan-Meier (K-M) curves of high and low HES4 expression groups. D K-M curve of high and low TGFBI expression groups.



**Figure 5.** Independent prognostic value of HES4. Forest plot of **A** univariate and **B** multivariate Cox analyses to identify independent prognostic factors in TCGA-BLCA samples. **C** Nomogram based on HES4 expression, stage, and age. **D** Nomogram calibration curve. **E** Decision curve analysis of the nomogram and independent prognostic factors.



**Figure 6.** HES4 function. Gene Set Enrichment Analysis (GSEA) diagram of HES4 based on **A** GO and **B** KEGG databases. Tree maps for the enriched **C** KEGG and **D** GO terms of 100 genes which were significantly co-expressed with HES4.



### 3.7 Differences in the immune microenvironment between high and low expression groups

The immune score, stromal score, ESTIMATE score, and tumor score were calculated for the high and low expression groups. Only the stromal score was significantly different between the two groups, with the low expression group having a higher stromal score (**Figure 7A**). The correlation between the four scores was explored, and it was found that the tumor purity was significantly negatively correlated with other indicators (stromal score:  $\text{cor} = -0.94$ ,  $P = 1.17 \times 10^{-170}$ ; immune score:  $\text{cor} = -0.92$ ,  $P = 2.08 \times 10^{-165}$ ; ESTIMATE score:  $\text{cor} = -1.0$ ,  $P = 0$ ), while the correlation between the remaining indicators was significantly positive (correlation of stromal score and immune score was  $0.74$ ,  $P = 2.21 \times 10^{-71}$ ; correlation of stromal score and ESTIMATE score was  $0.94$ ,  $P = 1.17 \times 10^{-190}$ ; correlation of immune score and ESTIMATE score was  $0.92$ ,  $P = 2.08 \times 10^{-165}$ ; **Figure 7B**). In total, 7 different algorithms were used to evaluate 119 immune cell differences between the high and low expression groups. A total of 33 immune cell types had significantly different abundances between the two groups, with T-cells, monocytes, macrophage/monocyte and neutrophils cells having a higher abundance (**Figure 7C**). The abundance of 8 cell types was significantly positively correlated with *HES4* expression, including central memory CD4+ T cells and NK T-cells, and the abundance of 18 cell types was significantly negatively correlated with *HES4* expression, including macrophages and mast cell (**Figure 7D**). TIDE analysis showed no significant differences between the high and low expression groups (**Supplementary Figure 5A**), and SubMap analysis indicated that the high *HES4* expression group was likely to be more sensitive to PD-1 therapy (**Figure 7E**). Nine further immune checkpoint molecules were not significantly different between the *HES4* expression groups (**Supplementary Figure 5B**).

### 3.8 Predicted AZD6482 sensitivity was correlated with *HES4* expression

The  $\text{IC}_{50}$  values of 138 drugs were predicted in the two *HES4* expression groups. 27 drugs had significantly different  $\text{IC}_{50}$  values between the two groups, with most drugs having a higher  $\text{IC}_{50}$  in the low expression group (**Supplementary Figure 6**). The correlation between *HES4* expression and the  $\text{IC}_{50}$  values of these 27 drugs was investigated, with 8 drugs having an  $\text{IC}_{50}$  value positively correlated with *HES4* expression, including AZD6482 and KIN001.135, and 19 drugs having an  $\text{IC}_{50}$  value negatively correlated with *HES4* expression, including NVP.TAE684 and CI.1040 (**Figure 8A**). To

further understand the potential therapeutic targets of the drug, AZD6482, which greatest correlation and significant difference between the high and low expression groups, was selected and docked to the *HES4* structure. As shown in **Figure 8B**, in the lowest energy docked structure, LYS-8 and SER-10 residues had hydrogen bonding interactions with AZD6482. The docking energy between AZD6482 and *HES4* was  $-4.0$  kcal/mol.

### 3.9 Assessment of mutations in the high and low *HES4* expression groups

Missense, frame shift deletion, splice site, frame shift insertion, nonsense, in-frame deletion, and multi-hit mutations were detected in the BLCA genomic data. The genes with the most mutations in both groups were *TP53* and *TTN*. Other genes were mutated at lower frequency, including *RYR2* and *RB1* (**Figure 9A-D**). The high expression group had higher TMB, indicating that the tumor cells in this group had higher neoantigen levels (**Figure 9E**). Finally, CNV was investigated. CNV location was displayed on chromosome maps (**Supplementary Figure 7A**), and both copy number amplification and deletion were strongly correlated with BLCA (**Supplementary Figure 7B-C**). CNV was significantly higher in the low expression group (**Supplementary Figure 7D**).

### 3.10 Regulatory and ceRNA network analysis of *HES4*

Four miRNAs were predicted to be associated with *HES4* using miRWalk, and 91 lncRNAs associated with 4 miRNAs were predicted using Starbase. Finally, *HES4*, the 4 miRNAs, and the 91 lncRNAs were used to construct a ceRNA network that contained 96 nodes and 119 edges (**Figure 10**).

### 3.11 Consistency of *HES4* expression trend in TCGA-BLCA and GSE121711

In both TCGA and GSE121711 datasets, *HES4* was up-regulated in cancer samples (TCGA:  $P < 0.01$ ; GSE121711:  $P < 0.05$ ; **Figure 11A-B**).

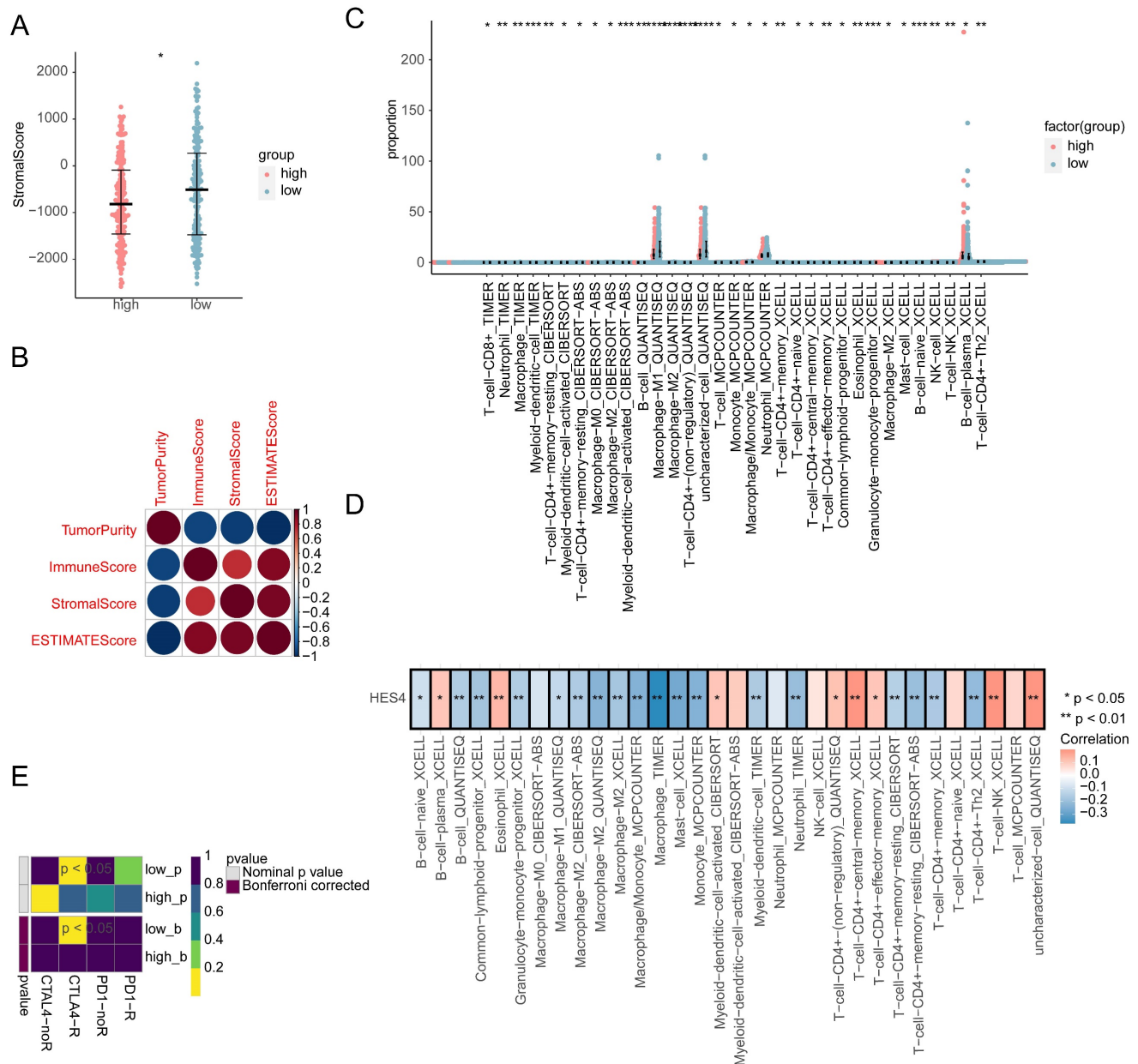
## 4. Discussion

Although though the majority of BLCA patients present with NMIBC at the first diagnosis, the prognosis progressively deteriorates on account of the highly recurrent and invasive nature of BLCA. Many genes associated with BLCA prognosis have been reported, along with genes whose expression and/or mutation can be used for diagnosis [30-32], but to date, there is still a lack of sufficiently sensitive and accurate molecular markers of BLCA to expedite diagnosis in order to both reduce the requirement for invasive investigations and to ensure timely provision

of treatment. In this study, we identified *HES4* as being causally associated with BLCA using transcriptomics and MR. *HES4* could both serve as an independent prognostic factor and play a crucial role in tumor progression, molecular interactions, and immune response processes.

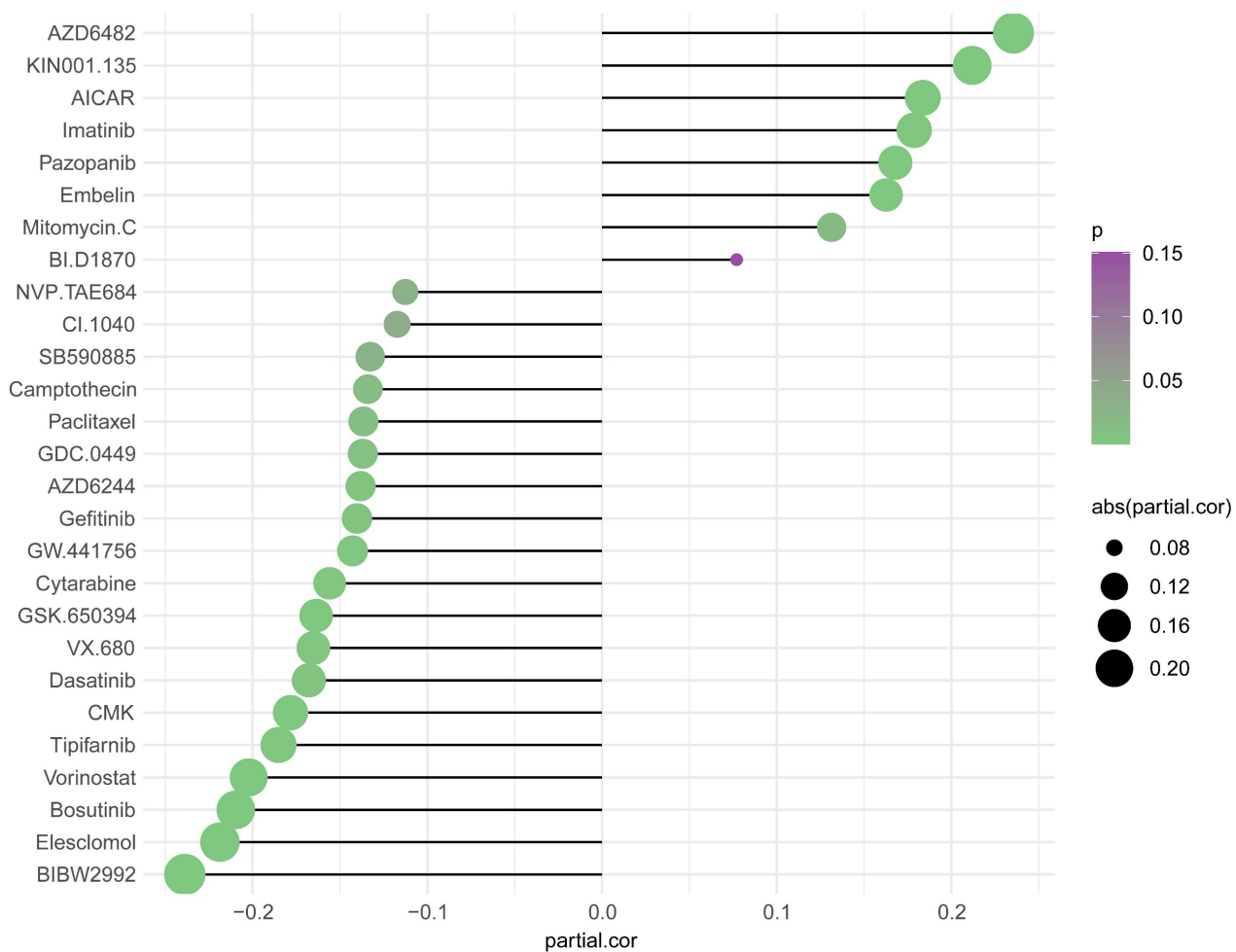
Following a similar approach, Li et al. [33] obtained a systemic lupus erythematosus (SLE)-related gene signature (SLEscore) for predicting the prognosis of patients with coexistent SLE and breast cancer, with 3- and 5-year AUCs of 0.81 and

0.91, respectively [34,35], indicating that a combined transcriptome/MR analysis can lead to more robust outcomes than transcriptome analysis alone. In the current study, we established a nomogram to predict the 1-, 3-, and 5-year survival rates of BLCA patients by combining age, stage, and *HES4* expression. Our results indicated this nomogram was most effective at predicting the 1-year survival rate of BLCA patients. The failure of the model to predict 3- and 5-year survival rates may imply that the development of BLCA is multifactorial.

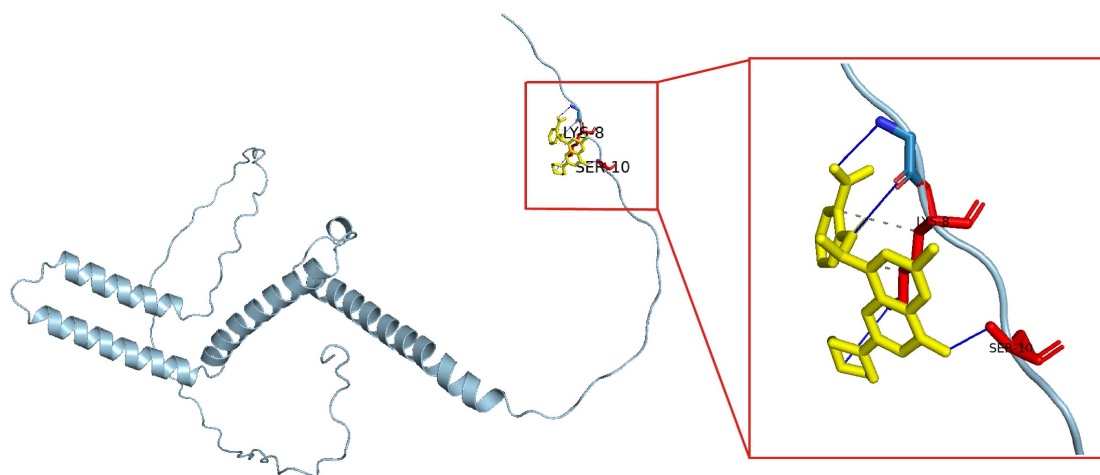


**Figure 7.** Immune infiltration analysis comparing the high and low *HES4* expression groups. **A** Stromal scores of the different expression groups. **B** Correlation heatmap of tumor purity, immune scores, stromal scores, and ESTIMATE scores. **C** Boxplots for the infiltration levels of 33 immune cells were calculated using different algorithms between high and low expression groups. **D** Correlation heatmap of differentially abundant immune cells and *HES4* expression. **E** SubMap analysis predicted that the high *HES4* expression group was more sensitive to PD-1 therapy.

A

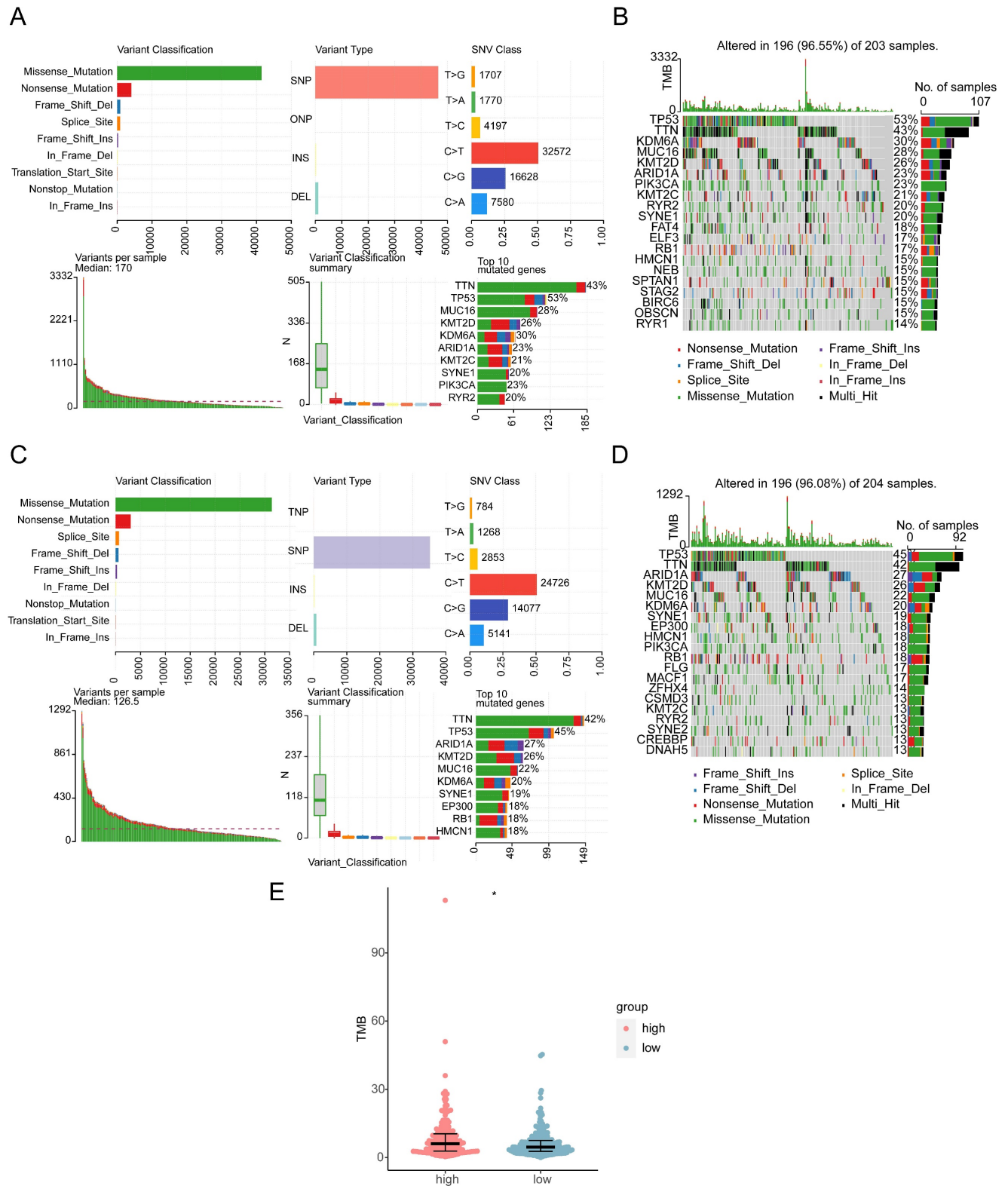


B

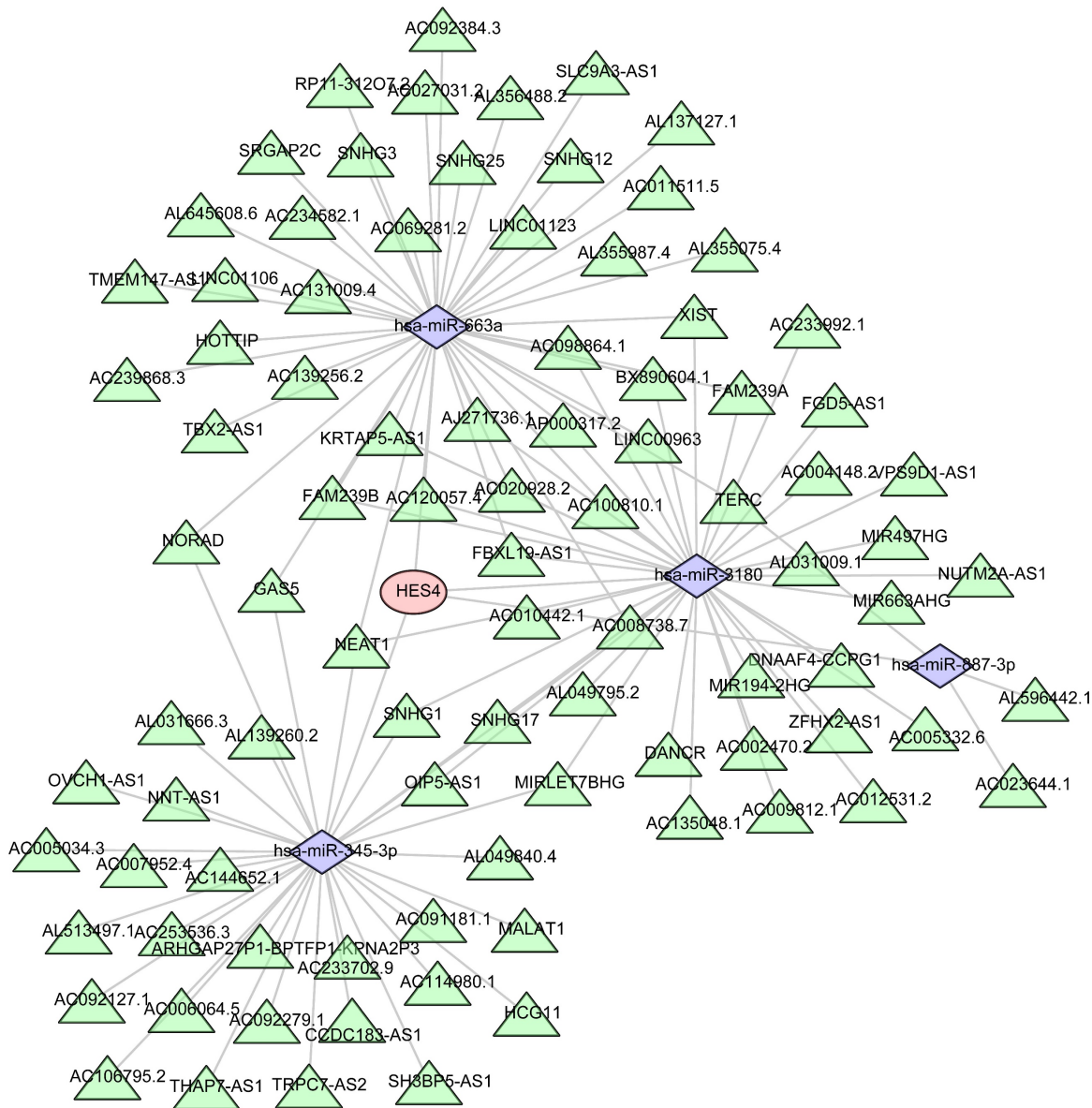


**Figure 8.** Drug sensitivity analysis in high and low *HES4* expression groups. **A** Correlation lollipop chart of the  $IC_{50}$  values of 27 drugs and *HES4* expression. **B** Molecular docking of AZD6482 to *HES4*.





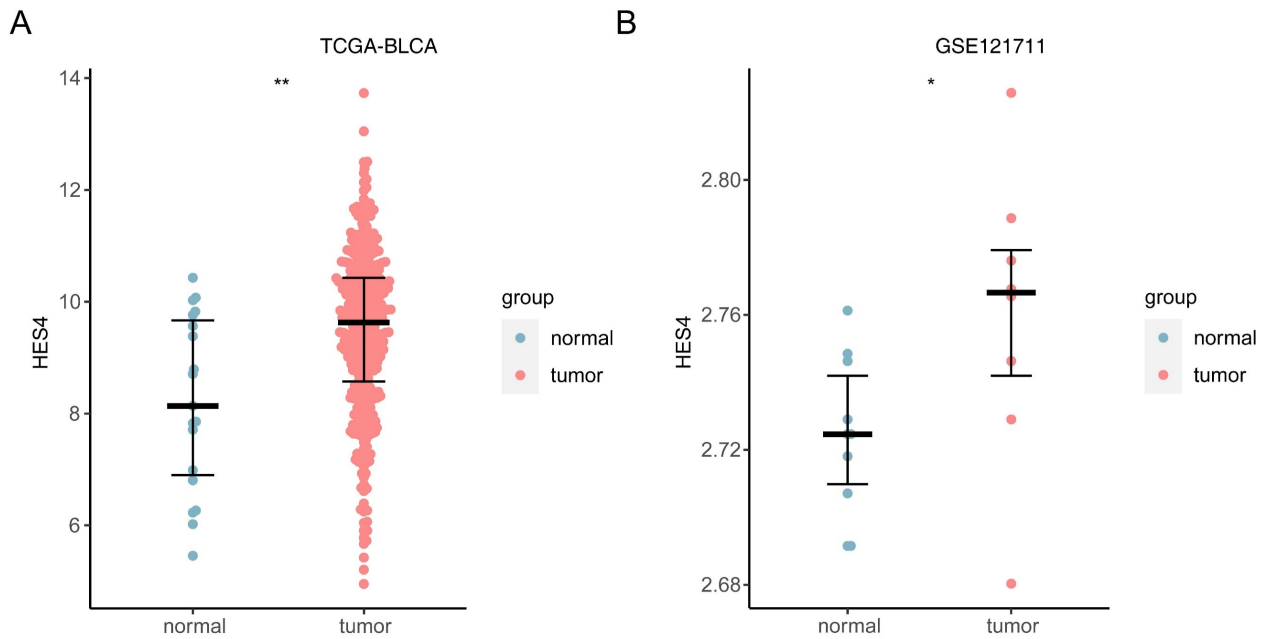
**Figure 9.** Mutational landscape and tumor mutation burden (TMB) in high and low *HES4* expression groups. **A** Overview of mutations in the high-risk subset of the TCGA-BLCA cohort. **B** Waterfall plot of eight specific mutation profiles of 203 patients with BLCA in the high-risk groups. The bar chart on the right shows the proportion of different mutation types of each gene, and the percentage indicates the mutation frequency of the gene. **C** Overview of mutations in the low-risk subset of the TCGA-BLCA cohort. **D** Waterfall plot of seven specific mutation profiles of 204 patients with BLCA in the low-risk groups. **E** Box plot of TMB in the different risk groups (Wilcoxon test, \* $p < 0.05$ ).



**Figure 10.** Prediction of the mRNA-miRNA-lncRNA regulatory network of *HES4*. Red represents *HES4*, purple represents miRNA, and green represents lncRNA.

*HES4*, a Hes family BHLH transcription factor, can repress the expression of target genes by recruiting co-repressors [36]. In the hematopoietic system, as a downstream gene of the Notch pathway, *HES4* promoted early T-lineage differentiation and induced T-cell development, while repressing the development of NK and myeloid cells [36]. Our immune infiltration analysis was consistent with this, with the expression level of *HES4* negatively correlated with macrophages/monocytes and NK cells, and positively correlated with some CD4<sup>+</sup> T cells. During tumorigenesis, high expression of *HES4* can increase the invasiveness of osteosarcoma cells and inhibit calcium deposition, and it is negatively correlated with OS [37]. High expression of the *HES4* gene was also reported to be significantly correlated

with activating mutations to Notch genes in breast cancers, which resulted in poorer prognosis [38]. In our study, the expression of *HES4* was significantly increased in BLCA patients, and the TMB level was higher in the group with high expression of *HES4*, yet the results of the multivariable cox regression analysis demonstrated that *HES4* may exert quite distinctive effects in different developmental stages of BLCA. GSEA did not identify enrichment of the Notch pathway. However, several other pathways were enriched, including pathways related to cellular senescence, energy metabolism, and neuronal diseases, suggesting that there may be additional important regulatory mechanisms beyond the Notch pathway in BLCA, which should be investigated in future experimental studies.



**Figure 11.** Expression of *HES4* in public datasets. **A** TCGA-BLCA. **B** GSE121711. \* $p < 0.05$ , \*\* $p < 0.01$ .

**Table 4.** Results for independent prognostic multivariate COX analysis.

id	coef	HR	HR.95L	HR.95H	pvalue
age	0.033528694	1.034097116	1.016531265	1.051966507	1.25E-04
STAGE	0.579303126	1.78479422	1.430066689	2.227511789	2.99E-07
<i>HES4</i>	-0.18997349	0.826981057	0.74121574	0.922670191	6.72E-04

AZD6482 is a phosphoinositide 3-kinase  $\beta$  (PI3K $\beta$ ) inhibitor [39]. The PI3K/Akt pathway and the MAPK pathway extensively interact in BLCA [40], low expression of *HES4* was associated with the MAPK pathway in GSEA analysis, and there was a positive correlation between the expression of *HES4* and the IC<sub>50</sub> value of AZD6482 in the GDSC database. Overall, these data suggest that the BLCA patients with low *HES4* expression may be more sensitive to AZD6482 treatment.

Numerous studies have reported that the tumor microenvironment plays an integral role in tumor progression, including in BLCA [41–43]. Parizi et al. demonstrated that tumor-infiltrating immune cells were strongly associated with the survival and biological behavior of BLCA patients [44]. To explore the immune microenvironment in BLCA and its relationship with *HES4*, we performed an immune infiltration analysis. The proportions of T-cells, monocytes, macrophages/monocytes, and neutrophils not only differed significantly between BLCA and normal samples, but were also strongly correlated with *HES4*. It has been reported that *ETV4* promotes lymphangiogenesis and lymph node metastasis in BLCA by regulating tumor-associated neutrophil infiltration; lymph node metastasis is the main cause of death in BLCA [45]. Huang et al. demonstrated that

*HSF1* was positively correlated with BLCA lymphatic metastasis and poor prognosis, and increased the infiltration of tumor-associated macrophages, especially M2 macrophages. We believe that the differences in immune cell infiltration in BLCA may be related to *HES4*, but the molecular mechanisms involved need to be explored in further experiments.

To our knowledge, this is the first study to use transcriptomics and MR analysis to identify a possible causal relationship between *HES4* expression and the occurrence and progression of BLCA. We constructed a nomogram to predict the 1-year survival of patients combining age, stage, and *HES4* expression. We preliminarily unveiled the potential complex mechanism of *HES4* in BLCA using enrichment analysis, immune infiltration analysis, and construction of a ceRNA network. We used the GDSC database and molecular docking to identify AZD6482 as a drug potentially targeting *HES4*, providing a novel candidate drug for the treatment of BLCA. This has important clinical implications for the early diagnosis, treatment, and prognosis of BLCA.

There are some limitations to our study. Firstly, this is an observational study and causality cannot be confirmed. Although MR analysis can provide some evidence of causality, experimental validation is still required. Secondly, although we used transcriptome



analysis to investigate the mechanisms by which *HES4* might promote BLCA, further experimental studies are needed to verify these mechanisms. Finally, although a potential target drug was identified, its efficacy and safety still need to be validated in animal models and clinical trials.

Overall, this study provides new clues to our understanding of the relationship between *HES4* and BLCA and offers new directions for the treatment of BLCA.

## Conclusions

*HES4* was identified as a diagnostic and prognostic biomarker for BLCA, and a nomogram combining age, stage, and *HES4* expression predicts the 1-year survival of patients. We also identified AZD6482 as a potential drug for targeted treatment of BLCA, which may be significant for clinical decision-making in BLCA patients.

## Supplementary Material

Supplementary figures and tables.

<https://www.jcancer.org/v15p1624s1.zip>

## Acknowledgments

### Funding

This work was supported by the National Natural Science Foundation of China (No. 82060464), the Yunnan Provincial Department of Science and Technology (Kun Medical Joint Special) Science and Technology Plan Project (202001AY070001-061), and the Yunnan Provincial Education Department Scientific Research Fund Project (2020J0187).

### Consent for publication

All authors consent to publication.

### Author contributions

Zhiyong Tan, Xiaorong Chen, Haihao Li, and Yinglong Huang contributed equally to the study's inception and design. Shi Fu, Mingxia Ding, Haifeng Wang and Jiansong Wang helped to revise the manuscript. All authors have read and approved the final manuscript.

### Data availability statement

1. The TCGA data are openly available at <https://portal.gdc.cancer.gov/>.

2. The GEO data are openly available at <https://www.ncbi.nlm.nih.gov/geo/>.

### Competing Interests

The authors have declared that no competing interest exists.

## References

- Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer statistics, 2022. *CA Cancer J Clin.* 2022;72(1):7-33.
- Jubber I, Ong S, Bukavina L, et al. Epidemiology of Bladder Cancer in 2023: A Systematic Review of Risk Factors. *Eur Urol.* 2023;84(2):176-190.
- Babjuk M, Burger M, Capoun O, 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.
- Magers MJ, Lopez-Beltran A, Montironi R, et al. Staging of bladder cancer. *Histopathology.* 2019;74(1):112-134.
- Witjes JA, Bruins HM, Cathomas R, et al. European Association of Urology Guidelines on Muscle-invasive and Metastatic Bladder Cancer: Summary of the 2020 Guidelines. *Eur Urol.* 2021;79(1):82-104.
- Patel VG, Oh WK, Galsky MD. Treatment of muscle-invasive and advanced bladder cancer in 2020. *CA Cancer J Clin.* 2020;70(5):404-423.
- Chen X, Zhang J, Ruan W, et al. Urine DNA methylation assay enables early detection and recurrence monitoring for bladder cancer. *J Clin Invest.* 2020;130(12):6278-6289.
- Zhou D, Jiang Y, Zhong X, et al. A unified framework for joint-tissue transcriptome-wide association and Mendelian randomization analysis. *Nat Genet.* 2020;52(11):1239-1246.
- Bowden J, Holmes MV. Meta-analysis and Mendelian randomization: A review. *Res Synth Methods.* 2019;10(4):486-496.
- Davey Smith G, Hemani G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Hum Mol Genet.* 2014;23(R1):R89-98.
- Burgess S, Thompson SG. Multivariable Mendelian randomization: the use of pleiotropic genetic variants to estimate causal effects. *Am J Epidemiol.* 2015;181(4):251-260.
- Yuan S, Larsson SC. Coffee and Caffeine Consumption and Risk of Kidney Stones: A Mendelian Randomization Study. *Am J Kidney Dis.* 2022;79(1):9-14.e11.
- Deng Y, Wu T, Luo G, Chen L. Exploring the casual association between coffee intake and bladder cancer risk using Mendelian Randomization. *Front Genet.* 2022;13:992599.
- Sun JX, Xu JZ, Liu CQ, et al. The association between human papillomavirus and bladder cancer: Evidence from meta-analysis and two-sample mendelian randomization. *J Med Virol.* 2023;95(1):e28208.
- Xiong J, Yang L, Deng YQ, et al. The causal association between smoking, alcohol consumption and risk of bladder cancer: A univariable and multivariable Mendelian randomization study. *Int J Cancer.* 2022;151(12):2136-2143.
- Du W, Wang T, Zhang W, Xiao Y, Wang X. Genetically supported causality between benign prostate hyperplasia and urinary bladder neoplasms: A mendelian randomization study. *Front Genet.* 2022;13:1016696.
- Babjuk M, Burger M, Compérat EM, et al. European Association of Urology Guidelines on Non-muscle-invasive Bladder Cancer (TaT1 and Carcinoma In Situ) - 2019 Update. *Eur Urol.* 2019;76(5):639-657.
- McDermaid A, Monier B, Zhao J, Liu B, Ma Q. Interpretation of differential gene expression results of RNA-seq data: review and integration. *Brief Bioinform.* 2019;20(6):2044-2054.
- Zhou X, Du J, Liu C, et al. A Pan-Cancer Analysis of CD161, a Potential New Immune Checkpoint. *Front Immunol.* 2021;12:688215.
- Zhou H, Zhang Y, Liu J, et al. Education and lung cancer: a Mendelian randomization study. *Int J Epidemiol.* 2019;48(3):743-750.
- Luo J, Xu Z, Noordam R, van Heemst D, Li-Gao R. Depression and Inflammatory Bowel Disease: A Bidirectional Two-sample Mendelian Randomization Study. *J Crohns Colitis.* 2022;16(4):633-642.
- Xu J, Zhang S, Tian Y, et al. Genetic Causal Association between Iron Status and Osteoarthritis: A Two-Sample Mendelian Randomization. *Nutrients.* 2022;14(18).
- Wang K, Guan C, Shang X, et al. A bioinformatic analysis: the overexpression and clinical significance of FCGBP in ovarian cancer. *Aging (Albany NY).* 2021;13(5):7416-7429.
- Scire J, Huisman JS, Grosu A, et al. estimateR: an R package to estimate and monitor the effective reproductive number. *BMC Bioinformatics.* 2023;24(1):310.
- Jiang P, Gu S, Pan D, et al. Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response. *Nat Med.* 2018;24(10):1550-1558.
- Prestegard JH. A perspective on the PDB's impact on the field of glycobiology. *J Biol Chem.* 2021;296:100556.
- Kim S, Chen J, Cheng T, et al. PubChem in 2021: new data content and improved web interfaces. *Nucleic Acids Res.* 2021;49(D1):D1388-d1395.
- Chen ZA, Tian H, Yao DM, et al. Identification of a Ferroptosis-Related Signature Model Including mRNAs and lncRNAs for Predicting Prognosis and Immune Activity in Hepatocellular Carcinoma. *Front Oncol.* 2021;11:738477.
- Mayakonda A, Lin DC, Assenov Y, Plass C, Koeffler HP. Maftools: efficient and comprehensive analysis of somatic variants in cancer. *Genome Res.* 2018;28(11):1747-1756.
- Ng K, Stenzl A, Sharma A, Vasdev N. Urinary biomarkers in bladder cancer: A review of the current landscape and future directions. *Urol Oncol.* 2021;39(1):41-51.

31. Lindskog SV, Prip F, Lamy P, et al. An integrated multi-omics analysis identifies prognostic molecular subtypes of non-muscle-invasive bladder cancer. *Nat Commun.* 2021;12(1):2301.
32. Hu J, Yu A, Othmane B, et al. Siglec15 shapes a non-inflamed tumor microenvironment and predicts the molecular subtype in bladder cancer. *Theranostics.* 2021;11(7):3089-3108.
33. Li W, Wang R, Wang W. Exploring the causality and pathogenesis of systemic lupus erythematosus in breast cancer based on Mendelian randomization and transcriptome data analyses. *Front Immunol.* 2022;13:1029884.
34. Liu Z, Mi M, Li X, et al. A lncRNA prognostic signature associated with immune infiltration and tumour mutation burden in breast cancer. *J Cell Mol Med.* 2020;24(21):12444-12456.
35. Yang X, Weng X, Yang Y, et al. A combined hypoxia and immune gene signature for predicting survival and risk stratification in triple-negative breast cancer. *Aging (Albany NY).* 2021;13(15):19486-19509.
36. Kageyama R, Ohtsuka T, Kobayashi T. The Hes gene family: repressors and oscillators that orchestrate embryogenesis. *Development.* 2007;134(7):1243-1251.
37. McManus M, Kleinerman E, Yang Y, et al. Hes4: A potential prognostic biomarker for newly diagnosed patients with high-grade osteosarcoma. *Pediatr Blood Cancer.* 2017;64(5).
38. Stoeck A, Lejnine S, Truong A, et al. Discovery of biomarkers predictive of GSI response in triple-negative breast cancer and adenoid cystic carcinoma. *Cancer Discov.* 2014;4(10):1154-1167.
39. Nylander S, Kull B, Björkman JA, et al. Human target validation of phosphoinositide 3-kinase (PI3K) $\beta$ : effects on platelets and insulin sensitivity, using AZD6482 a novel PI3K $\beta$  inhibitor. *J Thromb Haemost.* 2012;10(10):2127-2136.
40. Zhi Y, Cai C, Xu T, et al. Silencing of FGF6 hampers aerobic glycolysis and angiogenesis in bladder cancer by regulating PI3K/Akt and MAPK signaling pathways. *J Biochem Mol Toxicol.* 2023;37(8):e23399.
41. Pitt JM, Marabelle A, Eggermont A, et al. Targeting the tumor microenvironment: removing obstruction to anticancer immune responses and immunotherapy. *Ann Oncol.* 2016;27(8):1482-1492.
42. Tan Z, Chen X, Zuo J, et al. Comprehensive analysis of scRNA-Seq and bulk RNA-Seq reveals dynamic changes in the tumor immune microenvironment of bladder cancer and establishes a prognostic model. *J Transl Med.* 2023;21(1):223.
43. Liang S, Fang K, Li S, Liu D, Yi Q. Immune Microenvironment Terms Signature Robustly Predicts the Prognosis and Immunotherapy Response in Bladder Cancer Based on Large Population Cohorts. *Front Genet.* 2022;13:872441.
44. Kardoust Parizi M, Shariat SF, Margulis V, Mori K, Lotan Y. Value of tumour-infiltrating immune cells in predicting response to intravesical BCG in patients with non-muscle-invasive bladder cancer: a systematic review and meta-analysis. *BJU Int.* 2021;127(6):617-625.
45. Zhang Q, Liu S, Wang H, et al. ETV4 Mediated Tumor-Associated Neutrophil Infiltration Facilitates Lymphangiogenesis and Lymphatic Metastasis of Bladder Cancer. *Adv Sci (Weinh).* 2023;10(11):e2205613.