Molecular Diagnosis & Therapy (2021) 25:629–646
https://doi.org/10.1007/s40291-021-00539-1

ORIGINAL RESEARCH ARTICLE

Prognostic Biomarker-Based Identification of Drugs for Managing the Treatment of Endometrial Cancer

Dilraj Kaur[1] · Chakit Arora[1] · Gajendra Pal Singh Raghava[1]

Accepted: 27 May 2021 / Published online: 22 June 2021

Abstract

Introduction Uterine corpus endometrial carcinoma (UCEC) causes thousands of deaths per year. To improve the overall survival of patients with UCEC, there is a need to identify prognostic biomarkers and potential drugs.

Objectives The aim of this study was twofold: the identification of prognostic gene signatures from expression profiles of pattern recognition receptor (PRR) genes and identification of the most effective existing drugs using the prognostic gene signature.

Methods This study was based on the expression profile of PRR genes of 541 patients with UCEC obtained from The Cancer Genome Atlas. Key prognostic signatures were identified using various approaches, including survival analysis, network, and clustering. Hub genes were identified by constructing a co-expression network. Representative genes were identified using *k*-means and *k*-medoids-based clustering. Univariate Cox proportional hazard (PH) analysis was used to identify survival-associated genes. 'cmap2' was used to identify potential drugs that can suppress/enhance the expression of prognostic genes.

Results Models were developed using hub genes and achieved a maximum hazard ratio (HR) of 1.37 ($p = 0.294$). Then, a clustering-based model was developed using seven genes (HR 9.14; $p = 1.49 \times 10^{-12}$). Finally, a nine gene-based risk stratification model was developed (*CLEC1B*, *CLEC3A*, *IRF7*, *CTSB*, *FCN1*, *RIPK2*, *NLRP10*, *NLRP9*, and *SARM1*) and achieved HR 10.70; $p = 1.1 \times 10^{-12}$. The performance of this model improved significantly in combination with the clinical stage and achieved HR 15.23; $p = 2.21 \times 10^{-7}$. We also developed a model for predicting high-risk patients (survival ≤ 4.3 years) and achieved an area under the receiver operating characteristic curve (AUROC) of 0.86.

Conclusion We identified potential immunotherapeutic agents based on prognostic gene signature: hexamethonium bromide and isoflupredone. Several novel candidate drugs were suggested, including human interferon-α-2b, paclitaxel, imiquimod, MESO-DAP1, and mifamurtide. These biomolecules and repurposed drugs may be utilised for prognosis and treatment for better survival.

1 Introduction

Uterine corpus endometrial carcinoma (UCEC) or endometrial cancer (EC) is the sixth most common malignancy in females. Of the 65,620 reported cases, 12,590 deaths were estimated in 2020 [1]. In contrast with the declining trends for many common cancers, mortality has remained roughly the same for EC [2]. Advances in high-throughput technology mean the prognosis and diagnosis of EC at an early stage is possible more often. Still, a notable proportion of

Key Points

We identified the prognostic role of pattern recognition receptor genes in uterine corpus endometrial carcinoma (UCEC).

Prognostic biomarkers were selected using the network/cluster approach.

A panel of nine genes was identified as the best prognostic biomarkers for UCEC.

High-risk patients were predicted with high precision using these biomarker genes.

We recommend potential repurposed drug candidates for the treatment of patients with UCEC at high risk.

✉ Gajendra Pal Singh Raghava
raghava@iiitd.ac.in
http://webs.iiitd.edu.in/raghava/

[1] Department of Computational Biology, Indraprastha Institute of Information Technology-Delhi, Okhla Industrial Estate, New Delhi 110020, India

patients who develop metastasis or recurrent tumour have unfavourable prognoses. UCEC is classified under two major subtypes: type I tumours represent about 75–80% of the pathologic subtypes and are endometrioid adenocarcinomas [3, 4]; type II tumours are serous carcinomas and are less well-differentiated so have a poorer prognosis [4, 5]. Poor diagnosis and prognostic factors lead to high mortality and recurrence in patients with UCEC. To date, a multistep diagnosis process including gynaecological examination, transvaginal ultrasonography, and endometrial biopsy is required to distinguish UCEC from other benign diseases. At the same time, clinical features such as tumour grade, cervical involvement, lymph node status, histological subtype, depth of myometrial invasion, and lymphovascular space invasion (LVSI) are used as prognostic factors for patients with UCEC [6]. Biopsies provide information related to tumour grade, histological subtype, and other clinical features. Although platinum-based chemotherapy and hormonal therapy are the first-line therapies for UCEC, standard methods of treating UCEC consist of primary hysterectomy and bilateral salpingo-oophorectomy [7]. However, the occurrence of distant metastases mean patients respond poorly to conventional therapies and have a very low 5-year survival rate of ~ 17% [8]. Thus, efficient risk stratification methods are required for prognostic evaluation and therapeutic decision making in UCEC.

With the advent of high-throughput sequencing methods and public databases, many biomarkers have been identified for UCEC diagnosis, classification, and prognosis. These biomarkers, in contrast to clinicopathological factors, are related to underlying molecular mechanisms and offer robust theories for the pathogenesis of EC. Previous studies revealed that a set of five genes (*BUB1B*, *CCNB1*, *CDC20*, *DLGAPS*, *NCAPG*) is helpful in the prognosis of endometrial I carcinoma. Expression of these genes was higher and helped in predicting a higher tumour grade and worse overall survival (OS) [9]. Nine genes associated with glycolysis (*CLDN9*, *AK4*, *PC*, *GPC1*, and *SRD5A3*) were associated with poorer survival, whereas *B4GALT1*, *GMPPB*, *B4GALT4*, and *CHST6* were associated with better survival. Amongst these genes, *GMPPB* had the highest hazard ratio (HR; 1.544) and a *p* value of 0.0134 [10]. Higher *KLHL14* expression was associated with worse OS ($p = 0.0370$) and progression-free survival ($p = 0.081$) in UCEC samples [11]. Six potential prognostic tumour microenvironment (TME)-related genes (*CACNA2D2*, *CTSW*, *NOL4*, *SIGLEC1*, *TMEM150B*, and *TRPM5*) showed a positive correlation with OS in patients with UCEC [8].

For many years, the pattern recognition receptors (PRRs) have been known to have a role in recognising microbial ligands and the subsequent activation of the immune system. Recent advancements in the field of bioinformatics led to the development of a specific database related to PRRs [12]. The recently updated PRRDB2.0 [13] covers the latest information about receptors and their associated ligands and suggest that ligands for the toll-like receptors (TLRs), a well-known family of PRRs, show anti-tumoral activities in several cancers through activation in tumour cells. This activation could trigger both pro- and anti-tumoral effects depending on the context [14]. Increased angiogenesis and survival, enhancement of tumour invasion, and resistance to apoptosis are some of the functions indicating that TLRs act as tumour promoters [15]. The TLR pathways are critical regulators in chemo-resistance, possibly through activated nuclear factor (NF)-κB, upregulated expression of the anti-apoptotic protein B-cell lymphoma-2 (Bcl-2) [16]. Expression of TLRs such as *TLR3, 4, 7*, and *9* correlates with poor differentiation, high proliferation, and advanced staging in oesophageal cancer [17]. In lung cancer, *TLR5* expression is associated with a good prognosis, whereas *TLR7* is associated with a poor diagnosis. Also, expression of *TLR4*, *5*, *8*, and *9* are higher in patients with lung cancer than in healthy tissue [18]. *TLR7* and *8* are overexpressed and stage dependent in pancreatic cancer compared with healthy pancreas tissue [19]. Over-expression of *TLR1* suggested better survival in pancreatic cancer (HR 0.68; 95% confidence interval [CI] 0.47–0.99; $p = 0.044$) [20]. Under-expression of *TLR9* is associated with poorer survival in renal cell carcinoma [21] and is elevated in progression of glioma [22], and its lower expression is helpful in prediction of disease-free survival in triple-negative breast cancer [23]. In epithelial ovarian cancer, high expression of *TLR4* and *MyD88* predicted poorer OS in patients with endometrial ovarian cancer (HR 2.1; 95% CI 1.1–3.8) [24]. Studies have revealed that the combination of the variant alleles of the two *TLR9* polymorphisms, rs5743836 and rs187084, might be associated with a decreased risk for EC (odds ratio [OR] 0.11; 95% CI 0.03–0.44; $p = 0.002$) [25]. *TLR4* also provides chemoresistance in ovarian cancer cells [26]. *TLR3* and *TLR4* expression were examined during the menstrual cycle, endometriosis, postmenopausal endometrium, and endometrial hyperplasia, and low expression of *TLR3* and 4 were associated with poor prognosis in UCEC [27].

In this study, we aimed to investigate the altered expression profile of PRR genes in the context of survival prediction in patients with UCEC. The identification of core biomarker genes with a significant correlation with survival can be helpful in the stratification of risk groups and prognosis. Subsequently, the identified biomarker genes can form a reliable basis for the exploration of novel therapeutic strategies in UCEC treatment. Our study involved various bioinformatics approaches, such as the network-based approach, Cox-proportional hazard (PH) survival analyses, and the clustering-based approach for detection of the

critical genes and thereby development of highly accurate risk-prediction models. We also conducted a comprehensive assessment of the prognostic significance of various clinicopathological features and examined the molecular mechanisms attributed to the identified genes to find appropriate drug molecules that could increase the survival of patients with UCEC.

2 Materials and Methods

2.1 Dataset and Pre-Processing

The dataset was initially taken from 'The Cancer Genome Atlas' (TCGA) using TCGA Assembler 2, which consisted of quantile normalised RNA seq expression values for 581 patients with UCEC. The dataset had information about OS and censoring for only 541 patients. The list of 331 PRR signalling pathway genes was taken from Gene Set Enrichment Analysis and HUGO Gene Nomenclature Committee. Gene expression data were available for only 308 PRR genes. Thus, the final dataset was reduced to 541 samples constituting RNA seq values for 308 PRR-related genes using house Python and R scripts. Table 1 in electronic supplementary material (ESM)-1 summarises the clinical, demographic, and pathologic features corresponding to the UCEC final dataset.

2.2 Survival Analysis

HRs and 95% CI were computed to predict the risks of death associated with high-risk and low-risk groups based on the length of OS. These were stratified based on appropriate cut-offs for various factors, using the univariate unadjusted Cox-PH regression models. Kaplan–Meier (KM) plots were used to compare the survival curves of high-risk and low-risk groups. Survival analyses of these datasets were performed using 'survival' and 'survminer' packages (V.2.42-6) in R (V.3.4.4, The R Foundation). Statistical significance between the survival curves was estimated using log-rank tests. Wald tests were performed to evaluate the importance of the explanatory variables used for HR calculations. Concordance index provided the strength of the predictive ability of the model [28]. p values < 0.05 were considered significant. Multivariate survival analysis based on Cox regression was employed to compare the relationship between various covariates.

2.3 Gene Co-Expression Network

A PRR genes co-expression network was constructed to identify interactions among the genes. To formulate the gene co-expression network used here, we applied Pearson's correlation coefficient (PCC) for each PRR gene pair using gene expression value, to calculate statistically significant

Table 1 results of univariate Cox regression with > median cut-off

Gene	HR	1/HR	*p* Value	*C*-index	95% CI	Log-rank (*p*)	*q* Value
CLEC1B	6.48	0.15	2.11E−06	0.58	2.99–14.04	1.01E−04	4.78E−08
CLEC3A	2.71	0.37	3.47E−03	0.58	1.39–5.28	7.16E−03	2.34E−03
MRC1	2.17	0.46	1.60E−02	0.60	1.16–4.09	1.23E−02	1.35E−02
IRF7	0.47	2.14	1.77E−02	0.59	0.25–0.88	1.43E−02	2.56E−02
CTSB	0.50	2.00	2.65E−02	0.61	0.27–0.92	2.31E−02	3.64E−02
FCN1	2.00	0.50	2.85E−02	0.56	1.08–3.72	2.47E−02	3.53E−02
RIPK2	0.50	2.00	2.86E−02	0.57	0.27–0.93	2.42E−02	4.62E−02
CLEC3B	0.51	1.94	3.35E−02	0.59	0.28–0.95	2.96E−02	4.05E−02
CLEC12B	1.86	0.54	3.82E−02	0.57	1.03–3.35	4.11E−02	4.02E−02
TLR4	0.53	1.89	3.88E−02	0.55	0.29–0.97	3.55E−02	3.82E−02
NLRP10	1.94	0.52	3.99E−02	0.57	1.03–3.65	5.00E−02	3.57E−02
NLRP9	0.53	1.89	4.15E−02	0.56	0.29–0.98	3.70E−02	3.04E−02
MAPKAPK2	0.53	1.88	4.35E−02	0.55	0.29–0.98	3.89E−02	2.36E−02
TNIP1	0.54	1.86	4.38E−02	0.56	0.29–0.98	4.02E−02	2.56E−02
SARM1	0.54	1.85	4.95E−02	0.55	0.29–1.00	4.54E−02	1.53E−02

Genes with HR > 1 are bad BPM; those with HR < 1 are GPM

BPM bad prognostic maker, *CI* confidence interval, *C-index* concordance index, *GPM* good prognostic marker, *HR* hazard ratio, *q value* false discovery rate-corrected *p* value

associations. The 'Igraph' package in *R* was used to create an edge list from the obtained matrix using correlation cut-off, |PCC| > 0.5. Then, a gene co-expression network was constructed using 'Cytoscape' 3.7.1. The network analyser from Cytoscape was used to evaluate the degree of each node.

2.4 *k*-Means and *k*-Medoids Clustering

k-Means unsupervised clustering was used to identify a pre-specified number (*k*) of representatives/centroids. The package 'sklearn' was used to partition all PRR genes into *k* clusters in which each PRR gene belongs to the respective cluster with the nearest mean. *k*-Means clustering was done using Euclidean distance between gene expressions of 308 PRR genes. After *k* clusters were formed, a representative gene from each cluster was chosen. This choice was made based on the lowest *p* value based on univariate survival analysis. Similarly, medoids from each cluster were selected as representatives. This is because, unlike centroids, medoids are always restricted to genes of the clusters. Further, we used the obtained representative genes for model development.

2.5 Multiple Gene-Based Models

2.5.1 Machine Learning-Based Regression Models

Regression models from the 'sklearn' package in Python were implemented to fit the gene expression values (independent variables) against the OS time (target variable). Various regressors, including Linear, Random forest, K-nearest neighbours, Ridge, Lasso, Lasso Lars, and Elastic Net, were used. The fitting and test evaluations were carried using a fivefold cross-validation scheme, as implemented in previous studies [29]. A combination of all five evaluated test datasets (predicted OS) was then used to classify the actual patient OS at median cut-off to estimate HRs, CIs, and *p* values. Patients with a predicted OS higher than the median were classified as 'low risk', and those with a predicted OS lower than the median were classified as 'high risk'. Hyperparameter optimisation and regularisation was achieved using the in-built function 'GridsearchCV'. Model performance is denoted using the standard parameters, root mean squared error and mean absolute error.

2.5.2 Prognostic Index

As implemented in Li et al. [30] and Wang et al. [31], the prognostic index (PI) for a set of *k* genes was evaluated as shown in Eq. (1):

$$\mathrm{PI} = S_k b_k g_k \tag{1}$$

where β represents the regression coefficient obtained for a gene *g*, as estimated from a univariate Cox regression, the PI for a different set of genes was used for stratifying risk groups, and standard metrics such as HR, *p* value, etc. were estimated. Patients with a PI greater than the median of the PI were categorised as high risk, whereas patients with a PI less than the median of the PI were classified as low risk.

2.5.3 Gene Voting-Based Model

Corresponding to an individual gene expression (median cut-off), a risk label of 'high risk' or 'low risk' was assigned to each patient. Thus, for n survival-associated genes, every patient was denoted by a vector of *n* risk labels. In the gene voting-based method, the patient was ultimately classified into one of the high-/low-risk categories based on the dominant 'label' (i.e. occurring more than *n*/2 times) in this vector. This is shown in Fig. 1 for $n = 5$. This was followed by an evaluation of standard metrics.

3 Results

3.1 Survival-Associated Pattern Recognition Receptor Genes

A univariate Cox-PH analysis was done for all 308 PRR genes using median expression cut-offs. For a given gene, patients with an expression value greater or lower than the median expression value of the gene were classified as high

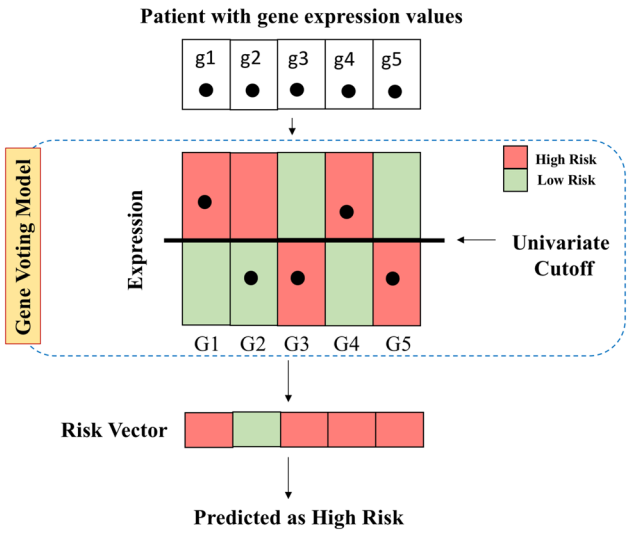

Fig. 1 Stratification of patients into high-/low-risk groups using five-gene voting model (G1, G2 … G5). The expression values of these genes (g1, g2, … g5) are used to determine the risk vector based on dominant risk labels

or low risk, respectively. The results of survival analysis in the form of HRs, concordance index, p values, CIs, and false discovery rate-corrected p values are provided in Table 2 in ESM 1. Of 308 genes, only 15 were significant ($p < 0.05$). The survival analysis revealed a total of nine good prognostic markers (GPM; i.e. genes positively correlated with patient OS time) and six bad prognostic markers (BPM; i.e. genes negatively correlated with patient OS time). Table 1 shows the results for these genes, along with the metrics associated with the stratification of high-/low-risk patients. The KM plots are shown in Fig. 1 in ESM 2. Table 3 in ESM 1 provides the precise molecular information about these 15 genes and the PubMed identifiers of the studies pertaining to their role in cancer, as obtained from Gene cards and The Candidate Cancer Gene Database [32], respectively. Table 4 in ESM 1 shows the results of risk stratification performed using various previously suggested prognostic genes in UCEC using Cox univariate analysis in the TCGA-UCEC dataset at median expression cut-off.

3.2 Risk Prediction Using Network-Based Features

In this section, we attempted to select features from the network of PRR genes. The aim of using the network was to identify features/representatives to understand connectivity in PRR genes. We used the following approaches for feature selections: (1) hub genes of the network, (2) medoids of clusters, and (3) representatives of clusters. These selected features or PRR genes were used to develop models to predict the survival of patients with cancer.

3.2.1 Hub Genes of Network

A correlation matrix was computed from 308 PRR genes where correlation among all possible pairs of genes was computed based on the expression provided in Table 5 in ESM 1. The correlation matrix used to create edges of the network using 'Igraph' for highly correlated pairs of genes ($|PCC| > 0.5$) as described in Sect. 2 is provided in Table 6 in ESM 1. We used 'Cytoscape' software to visualise and analyse the gene network. The effective correlation was set to be greater than 0.5, which resulted in 116 nodes and 804 edges. We selected the top 15 hub genes (*BTK*, *ITGB2*, *HAVCR2*, *FCRL3*, *CD163*, *CD300LF*, *CD68*, *CTSS*, *CLEC10A*, *CLEC12A*, *NR1H3*, *CLEC4E*, *CD209*, *ITGAM,* and *TLR8*) based on their degree. These hub genes were used to build a prognostic model for predicting the survival risk of patients with UCEC. Our voting-based model developed using these 15 hub genes achieved an HR of 1.37 with $p = 0.294$. The network is shown in Fig. 2, and the network metrics are provided in Table 7 in ESM 1.

3.2.2 Medoids of Clusters

We performed k-medoids clustering of genes based on pairwise dissimilarity. The medoids were selected for clusters at different $k = 5$, 10, 15, 20, and 25. The best result was obtained for the gene voting model at $k = 5$ (HR 1.85; $p = 0.045$). The results at different k values are provided in Table 8 in ESM 1.

3.2.3 Representative Clusters

The representative gene was chosen based on the lowest p value obtained from univariate survival analysis from each cluster. Obtained representative genes from each cluster were used to develop a risk stratification model. This process was repeated for $k = 5$, 10, 15, 20, and 25. The best result was obtained for $k = 10$ (HR 4.11; $p = 3.7 \times 10^{-5}$), as shown in Table 2. Further, we filtered the representative genes by using a cut-off of $p < 0.05$ within each cluster. The best results were obtained for gene voting models, as shown in Table 3. We observed that optimal risk segregation was obtained for $k = 15$, which resulted in seven representative genes. Table 9 in ESM 1 provides details about the distribution of 308 genes into 15 clusters. The gene voting model developed for the genes *CLEC1B*, *CLEC3A*, *CTSB*, *NLRP10*, *NLRP9*, *TNIP1*, and *SARM1* achieved HR 9.14 and $p = 1.49 \times 10^{-12}$. The network construction was done using 'Cytoscape 3.7.1' and shows seven different clusters with their representative genes in Fig. 3.

3.3 Risk Estimation Using Multiple Gene-Based Models

Several risk stratification models based on machine-learning-based regression (MLR), PI, and gene voting were constructed using the expression profile of survival-associated PRR genes (based on p value). We tried various combinations using 15 significant genes and found that a combination of nine genes performed the best: *CLEC1B*, *CLEC3A*, *IRF7*, *CTSB*, *FCN1*, *RIPK2*, *NLRP10*, *NLRP9*, and *SARM1*. Table 4 shows the results corresponding to various risk prediction models for these nine genes. Amongst these, the gene voting-based model performed the best: HR 10.70 and $p \sim 10^{-12}$. A concordance index value of 0.76 was also the highest for this model, and high-/low-risk group survival curves were significantly separated with a log-rank $p \sim 10^{-14}$. Figure 4 shows the KM plot representing the survival curves corresponding to the two risk groups. While the 5-year survival rate for low-risk patients was close to 85%, it dropped as low as 15% for high-risk patients. The PI-based model had the next-best performance, with an HR of 3.41 and $p \sim 10^{-3}$, and the regression-based linear model was the third

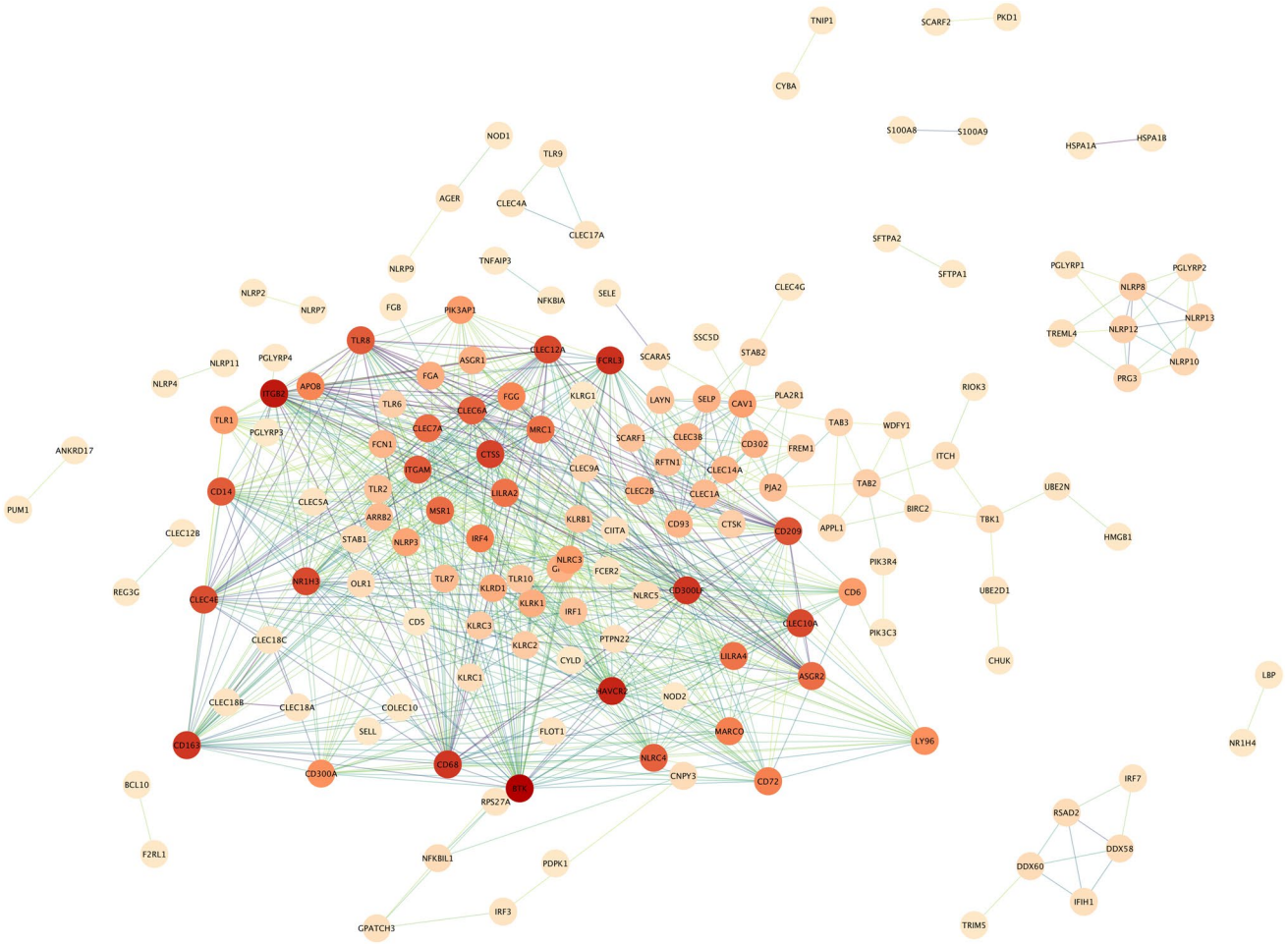

Fig. 2 Gene co-expression network analysis of pattern recognition receptor genes. Interconnection of 15 hub genes; darker colour represents a higher degree score, darker edge shows higher clustering coefficient value

Table 2 Results of gene voting model for chosen representative genes from each cluster

S.no	Genes	Clusters (*n*)	HR	*p* Value	*C*-index	95% CI	log-rank (*p*)
1	*CLEC1B, CTSB, NLRP10, UBC, LTF*	5	3.53	3.74E−05	0.64	1.94–6.43	6.84E−05
2	***CLEC1B, CTSB, NLRP10, CLEC3A, UBC, ESR1, LTF, UBB, LGALS3BP, MAPKAPK2***	**10**	**4.11**	**4.38E−06**	**0.64**	**2.25–7.51**	**1.57E−05**
3	*CLEC1B, UBC, LTF, CTSB, NLRP10, UBB, LGALS3BP, S100A9, CLEC3A, RPS27A, APPL1, HSPA1A, SARM1, NLRP9, TNIP1, S100A9*	15	2.93	8.00E−04	0.62	1.56–5.49	5.41E−04
4	*CLEC1B, UBC, LTF, CTSB, NLRP10, DMBT1, LGALS3BP, S100A9, CLEC3A, RPS27A, APPL1, HSPA1A, SARM1, NLRP9, TNIP1, HSPD1, UBB, CYBA, FLOT1, HMGB1*	20	3.64	2.06E−05	0.66	2.01–6.59	2.06E−05
5	*CLEC1B, UBC, LTF, CTSB, NLRP10, HSPA1A, LGALS3BP, S100A9, CLEC3A, RPS27A, BIRC3, MRC2, CNPY3, NLRP9, TNIP1, HSPD1, UBB, CYBA, MRC1, HMGB1, VCAN, CFI, S100A8, FCN1, SARM1*	25	3.58	4.87E−05	0.64	1.93–6.61	2.86E−05

Bold formatting indicates the set of representative genes for the best model

CI confidence interval, *C-index* concordance index, *HR* hazard ratio

Table 3 Results of gene voting model for chosen significant representative genes only from each cluster

S.no	Representative genes	Clusters (*n*)	HR	*p* Value	*C*-index	95% CI	Log-rank (*p*)
1	*CLEC1B, CTSB, NLRP10*	5	3.62	1.12E−04	0.60	1.88–6.96	4.87E−04
2	*CLEC1B, CTSB, NLRP10, CLEC3A*	10	4.25	1.63E−06	0.53	1.30–13.84	4.70E−02
3	***CLEC1B, CLEC3A, CTSB, NLRP10, NLRP9, TNIP1, SARM1***	**15**	**9.14**	**1.49E−12**	**0.73**	**4.95–16.87**	**6.64E−12**
4	*CLEC1B, CLEC3A, CTSB, NLRP10, NLRP9, TNIP1, SARM1*	20	9.14	1.49E−12	0.73	4.95–16.87	6.64E−12
5	*CLEC1B, CLEC3A, CTSB, NLRP10, NLRP9, TNIP1, SARM1, MRC1, FCN1*	25	5.46	3.05E−08	0.67	2.99–9.95	9.00E−08

Bold formatting indicates the set of significant representative genes for the best model

CI confidence interval, *C-index* concordance index, *HR* hazard ratio

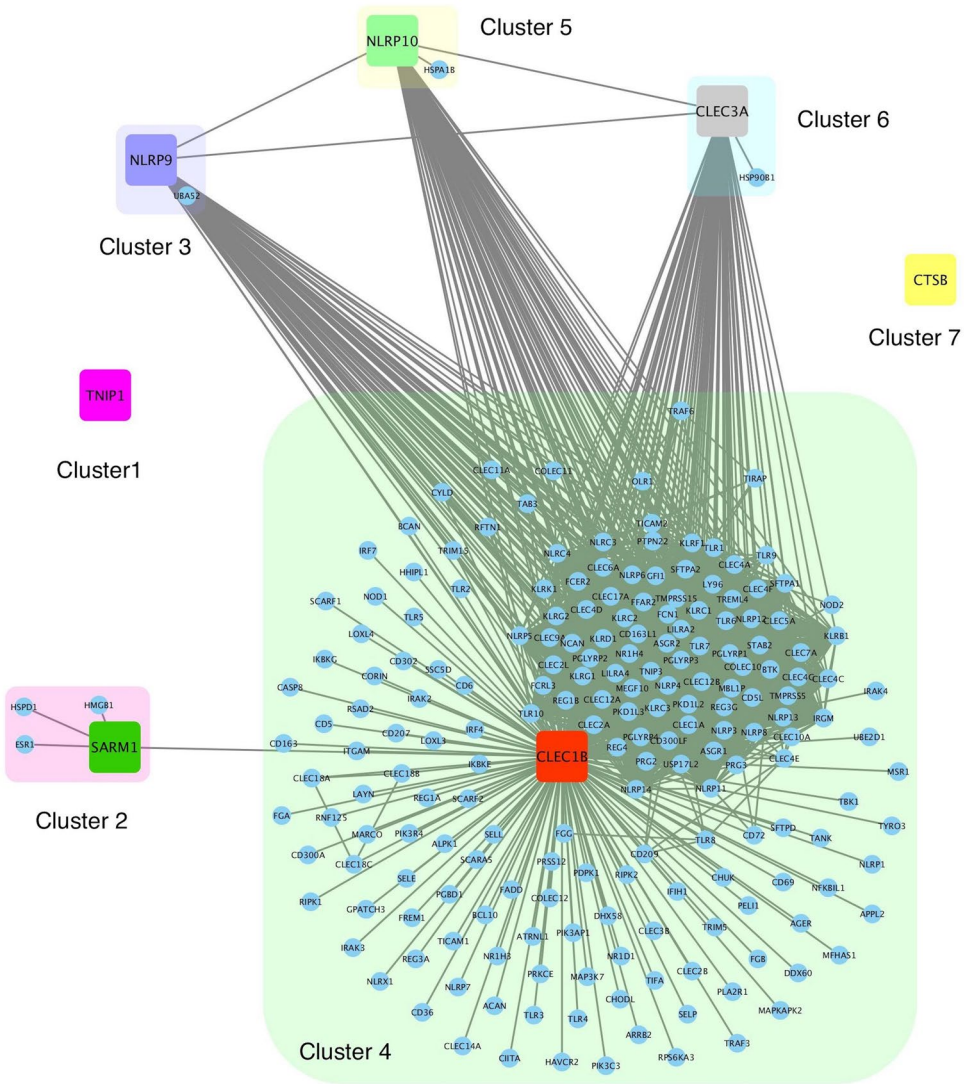

Fig. 3 Network based on clustering: seven different clusters are shown in different colours. Representative pattern recognition receptor genes are highlighted and represented within large squares

Table 4 Performance of different models developed using multiple gene expression profile

Model	HR	*p* Value	*C*-Index	95% CI	Log-rank (*p*)
Voting	10.70	1.13E−12	0.76	5.57–20.55	8.15E−14
PI	3.41	9.70E−03	0.60	1.35–8.66	2.59E−03
Linear	1.66	1.00E−01	0.56	0.91–3.03	9.81E−02
Ridge	0.99	9.86E−01	0.52	0.55–1.79	9.86E−01
KNN	0.83	5.24E−01	0.52	0.46–1.49	5.24E−01
Elastic net	0.79	4.45E−01	0.55	0.44–1.43	4.47E−01
Random forest	0.77	3.85E−01	0.55	0.43–1.39	3.86E−01
Lasso	0.65	1.57E−01	0.56	0.36–1.18	1.53E−01
SVR	0.65	1.57E−01	0.56	0.36–1.18	1.53E−01
Lasso Lars	0.65	1.57E−01	0.56	0.36–1.18	1.53E−01

Bold formatting indicates statistically significant results (*p* value, log-rank $p < 0.05$). Machine-learning-based regression hyperparameters are provided in Table 12 in the Electronic Supplementary Material

CI confidence interval, *C-index* concordance index, *HR* hazard ratio, *KNN* K-nearest neighbours, *PI* prognostic index, *SVR* support vector regression

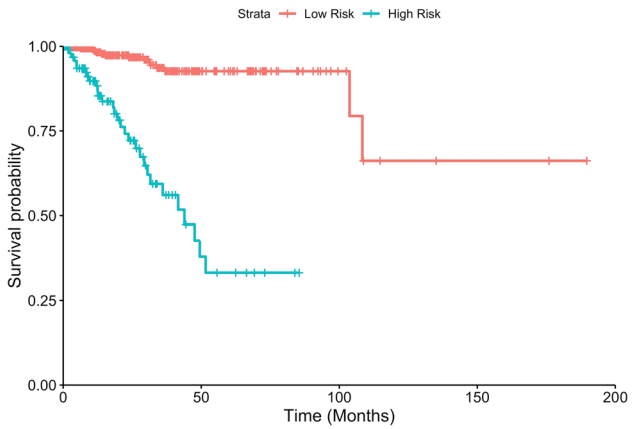

Fig. 4 Kaplan–Meier plot showing risk stratification of patients with uterine corpus endometrial carcinoma based on gene voting model. Patients with more than four 'high-risk' labels in the ten-bit risk vector are assigned (blue) as high risk (hazard ratio 10.70, $p = 1.13 \times 10^{-12}$, $C = 0.76$, log-rank-$p = 8.15 \times 10^{-14}$), whereas others were assigned low risk (red)

best (and top amongst the MLR models) with HR 1.66, but the *p* value was not statistically significant.

3.4 Multiple-Gene Model Sub-stratifies Patients in Clinicopathological High-Risk Groups

Previous studies indicated a role for clinicopathological factors in UCEC prognosis, such as histologic diagnosis, ethnicity, clinical stage [6], menopause status, peritoneal washing, etc. [33]. Therefore, we performed a univariate analysis to assess the association of these factors with OS in our dataset. Table 5 shows the results of the univariate analysis. Clinical factors such as clinical stage, residual tumour, peritoneal washing, grade, histologic grade, and menopause status were the significant factors in UCEC prognosis. The gene voting model was able to sub-stratify the high-risk patients with UCEC based on clinicopathological factors such as histologic diagnosis, peritoneal washing, menopause status, neoplasmic grade, residual tumour, and clinical stage, as shown in Fig. 5. The KM plots, along with low log-rank *p* values, denote a significant separation between high- and low-risk patients.

3.5 Multivariate Analysis

We performed a multivariate Cox regression survival analysis using seven prominent prognostic markers, i.e. multiple gene voting model, clinical staging, residual tumour, peritoneal washing, histological subtype, menopause status, grade. The p-value corresponding to the gene voting model (HR 8.17; $p < 0.001$) and the clinical stage (HR 3.11; $p = 0.03$) was significant, whereas it was not significant for others, as depicted by the forest plot in Fig. 6. Thus, a hybrid model can be made using the gene voting model and the clinical stage for further improvement in risk stratification.

3.6 Hybrid Voting Model

After obtaining the independent covariates, i.e. multiple gene voting model and the clinical stage, based on a multivariate Cox regression survival analysis, we developed a hybrid voting model. This model combined clinical stage with the nine gene voting models for risk stratification purposes. As such, the risk vector associated with each patient was a 10-bit vector, with 1 bit assigned to risk label because of the clinical stage. The model performed better than the nine-gene voting model (HR 15.23; $p = 2.21 \times 10^{-7}$, concordance index =0.78, log-rank-$p = 2.76 \times 10^{-17}$). Figure 7 shows the KM plot corresponding to the hybrid model.

3.7 Predictive Validation

As implemented in Zhao et al. [34], we performed a predictive assessment of our gene voting model using sub-samples of the complete dataset. Sampling sizes of 50%, 70%, and 90% were chosen with 100 iterations each. HR and concordance index were evaluated for each iteration corresponding to the gene voting model and the hybrid model. Figure 8 shows the boxplots corresponding to these results. The median HR (15.34, 15.32, 15.02) and concordance index

Table 5 Univariate analysis using clinicopathological features

Clinical factors	Strata	*N*	HR	*p* Value	*C*-Index	95% CI	Log-rank (*p*)
Clinical stage	III, IV vs. I, II	541.00	4.44	1.31E−06	0.71	2.43–8.11	8.81E−07
Residual tumour	R0 vs. R1, R2	411.00	0.29	8.06E−04	0.59	0.14–0.60	2.27E−03
Peritoneal washing	Negative vs. positive	407.00	0.31	2.42E−03	0.58	0.15–0.66	5.38E−03
Grade	G3, high grade vs. G1, G2	541.00	3.27	2.46E−03	0.61	1.52–7.03	7.24E−04
Histologic_diagnosis	MSE, SEA vs. EEA	541.00	2.29	6.61E−03	0.56	1.26–4.16	8.65E−03
Menopause status	Pre vs. post	512.00	2.54	3.51E−02	0.53	1.07–6.02	5.92E−02
Age	> 64 vs. ≤ 64	541.00	1.30	3.79E−01	0.54	0.72–2.33	3.79E−01
Race	White vs. others	511.00	1.15	6.83E−01	0.50	0.58–2.29	6.79E−01
Ethnicity	Hispanic or Latino vs. not Hispanic not Latino	388.00	1.44	7.23E−01	0.50	0.19–10.61	7.37E−01
History other malignancy	No vs. yes	541.00	1.18	7.84E−01	0.49	0.36–3.81	7.79E−01
Surgical approach	Open vs. minimally invasive	519.00	1.06	8.71E−01	0.51	0.54–2.07	8.71E−01

Bold formatting indicates statistically significant results (*p* value, log-rank $p < 0.05$). Clinical stage, residual tumour, peritoneal washing, grade, histologic washing, and menopause status are the significant factors

CI confidence interval, *C-index* concordance index, *EEA* endometrioid endometrial adenocarcinoma, *HR* hazard ratio, *MSE* mixed serous and endometrioid, *SEA* serous endometrial adenocarcinoma

(0.786, 0.792, 0.783) values for the hybrid model remained better despite the sampling size. This method ensured that the risk stratification models were robust and performed well with random datasets of different sizes.

3.8 Classification Using Hybrid Voting Model

We evaluated the performance of the hybrid model using the area under the receiver operating characteristic curve (AUROC) value. The 'survivalROC' package was used to calculate the true-positive and false-positive rates. Here, a prediction was considered a true positive if the OS was greater than the cut-off time and the patient was in a low-risk group according to the model; the converse applied for a true negative prediction, as shown in Fig. 9a. The median of the OS time was 1.1 years. We observed that, amongst different OS cut-offs, the model performed best at the cut-off of 4.3 years. This cut-off is a good classifier between high- and low-risk patients. Using this cut-off, an AUROC value of 0.86 was obtained by the classification based on the hybrid voting model. The ROC curve corresponding to this is shown in Fig. 9b.

3.9 Screening of Therapeutic Agents

The choice of therapy is a crucial step after the identification of genes that play a critical role. To do so, we extracted drug molecules that could re-modulate the overexpressed and underexpressed genes, using the 'Cmap2' database [35], as per Shen et al. [36]. We queried a list of probe identifiers corresponding to upregulated genes (*CLEC1B*, *CLEC3A*, *FCN1*, *NLRP10*) and downregulated genes (*RIPK2*, *SARM1*, *IRF7*, *CTSB*, *NLRP9*) as input to Cmap2. The output was ranked based on p-values (see Table 10 in ESM 1). The top positive and negative enriched molecules were hexamethonium bromide (enrichment 0.834; $p = 2.6 \times 10^{-4}$) and isoflupredone (enrichment − 0.955; $p = 2.2 \times 10^{-4}$), respectively. Hexamethonium bromide is a non-depolarising ganglion blocker and a nicotinic acetylcholine receptor antagonist. It can be used to treat hypertension and duodenal ulcers [37]. It is known to be poorly absorbed from the gastrointestinal tract and does not cross the blood–brain barrier. Isoflupredone, also known as delta-fludrocortisone and 9α-fluoroprednisolone, is a synthetic glucocorticoid corticosteroid. 3D conformers of these drug molecules obtained from PubChem [38] are shown in Fig. 10.

3.9.1 Repurposing of Drugs

We also assessed how approved drugs might be helpful in prolonging UCEC survival. We obtained five GPM and four BPM genes. The biological mechanisms of GPM genes (*RIPK2*, *SARM1*, *IRF7*, *CTSB*, *NLRP9*) that showed a good prognosis in UCEC can be exploited to identify positive regulators/inducers that can be used as potential therapeutics. Inducing *IRF7* can lead to the production of interferon (IFN)-α and IFNβ (Fig. 11a). These have a significant role in anti-tumour immunity and homeostasis and have been accepted by the US FDA to treat malignancies [39]. IFNα is known to trigger apoptosis in tumour cells by various stimuli and induce differentiation of dendritic cells [40]. In comparison, IFNβ stimulates immune cells, including macrophages and natural killer (NK) cells, that may boost the anti-tumour effect [41]. Human IFNα-2b is a synthetic replacement for

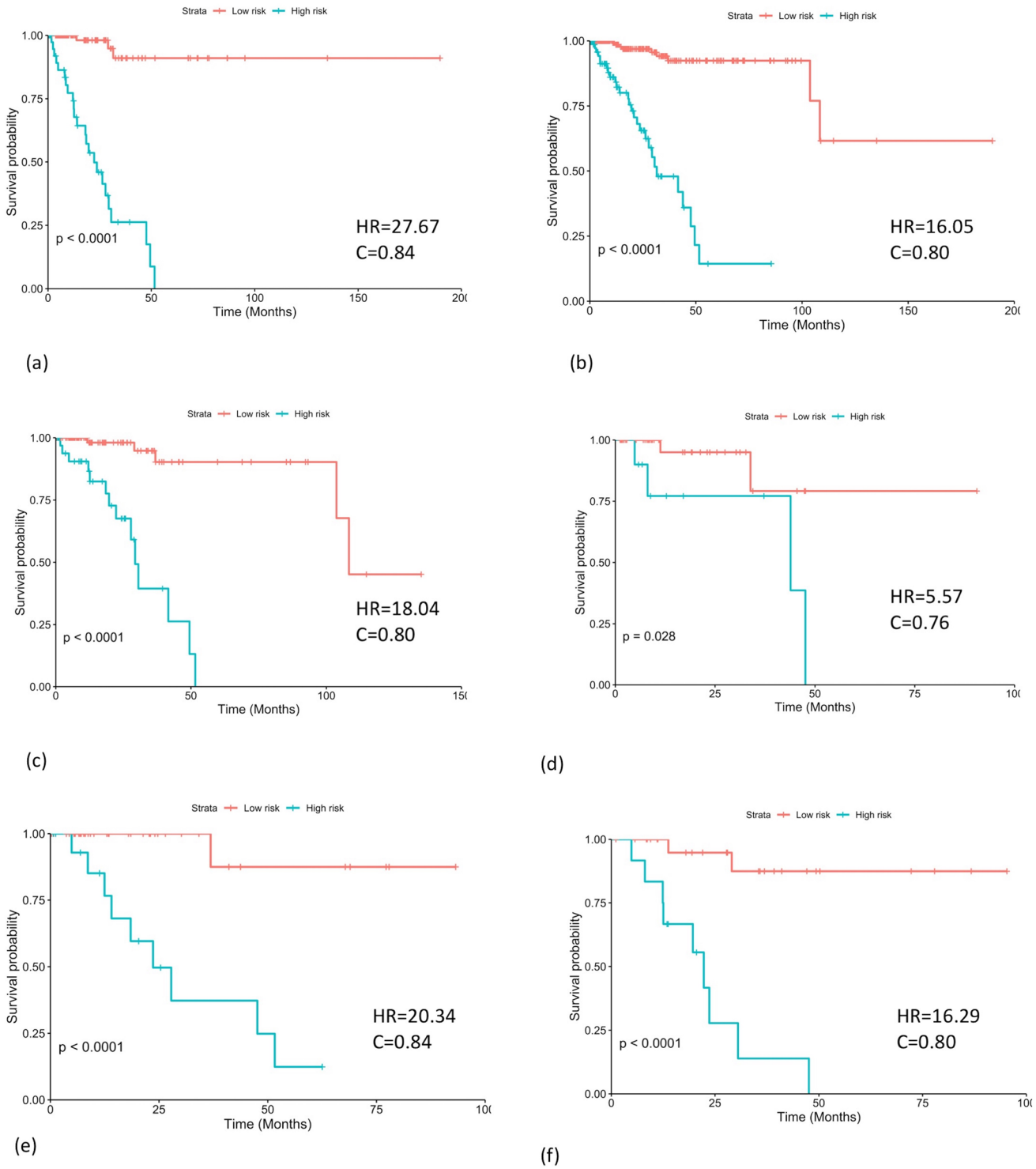

Fig. 5 Gene voting model sub-stratifies high-risk groups. **a** Patients with clinical stage III/IV (n = 152) were stratified into high- and low-risk groups with logrank-$p = 4.3 \times 10^{-13}$; **b** neoplasmic grade = G3 and high grade (n = 321) were stratified into high- and low-risk groups with logrank-$p = 5.0 \times 10^{-14}$; **c** histologic diagnosis = mixed serous adenocarcinoma, endometrioid endometrial adenocarcinoma (n = 136) were stratified into high- and low-risk groups with logrank-$p = 9.2 \times 10^{-8}$; **d** menopause status = pre (n = 52) were stratified into high- and low-risk groups with logrank-$p = 4.1 \times 10^{-9}$; **e** peritoneal washing = positive (n = 57) were stratified into high- and low-risk groups with logrank-$p = 8.4 \times 10^{-5}$; **f** residual tumour = R1, R2 (n = 38) were stratified into high- and low-risk groups with logrank-$p = 3.2 \times 10^{-5}$. *HR* Hazard Ratio and *C* Concordance index

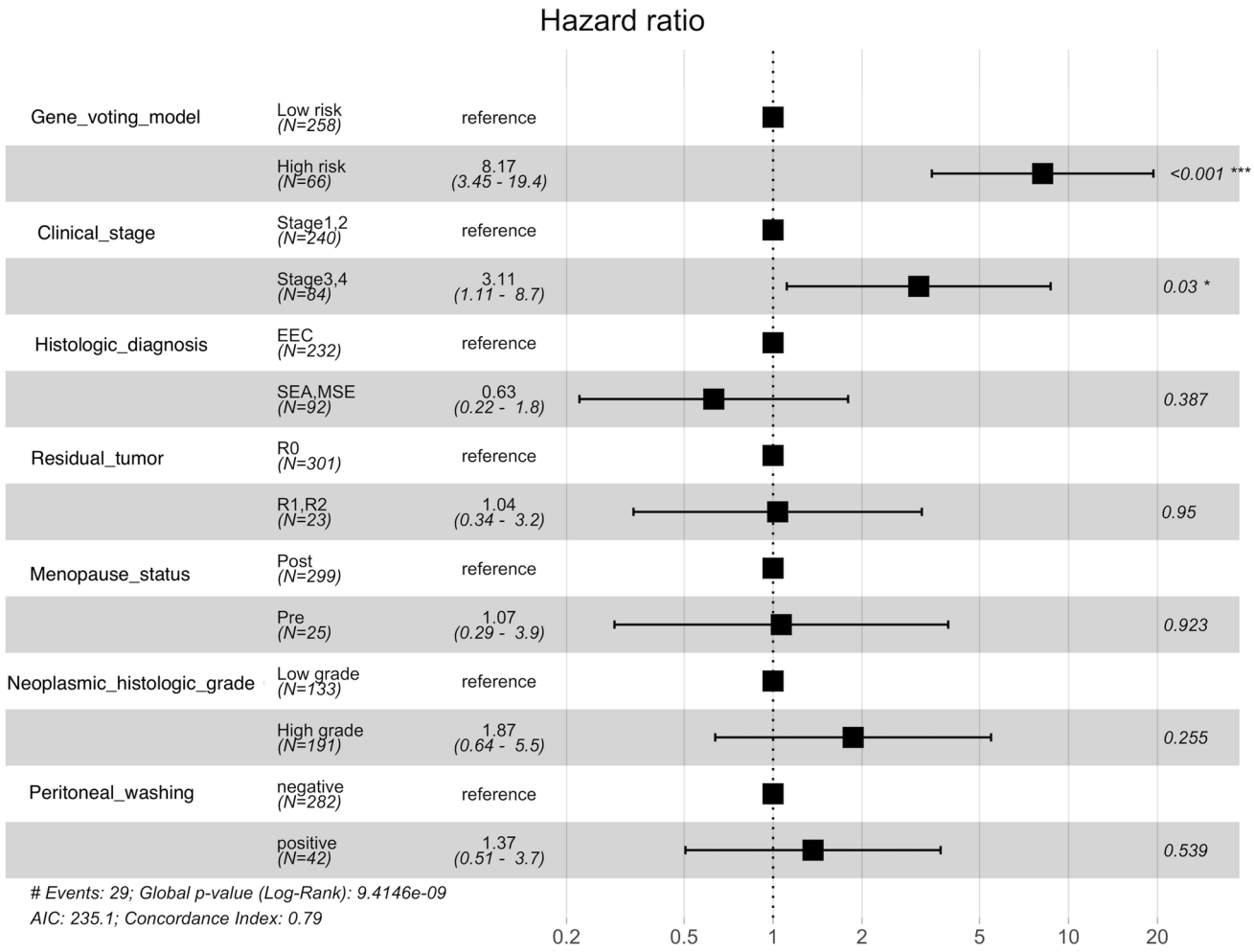

Fig. 6 Multivariate analysis reveals gene voting model (hazard ratio 8.17; $p < 0.001$) and clinical stage (hazard ratio 3.11; $p = 0.03$) as independent covariates

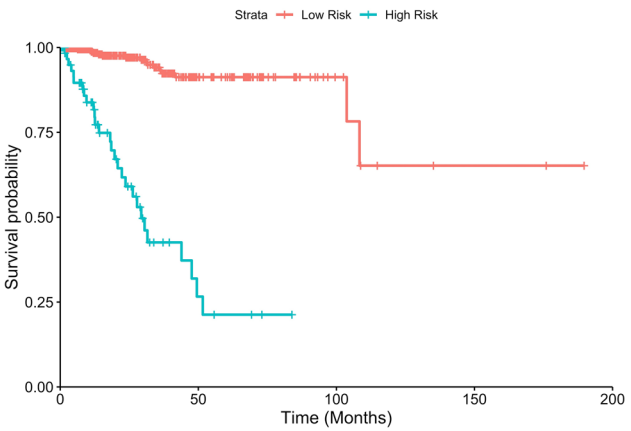

Fig. 7 Hybrid model for risk stratification using nine-gene voting model and clinical stage (hazard ratio 15.23; $p = 2.21 \times 10^{-7}$, $C = 0.78$, log-rank-$p = 2.76 \times 10^{-17}$)

IFNα and is approved by the FDA as an immunomodulatory drug. It has been used in hairy cell leukaemia [40], follicular lymphoma [42], and renal cell carcinoma [43]. Paclitaxel and imiquimod are the positive regulators for *TLR4* and *TLR7*, respectively [44, 45], and can be used as therapeutic drugs for indirect induction of *IRF7* and thereby IFNs. Agonists such as lipopolysaccharide and imidazoquinoline anti-viral compound single strand RNA may also help to induce the signalling pathway and lead to the production of IFNα/β. *CTSB*, which plays a crucial role in the NLRP signalling pathway can be positively regulated through various agonists such as pathogen-associated molecular patterns/damage-associated molecular patterns (Fig. 11b). Different interleukin (IL) therapeutics may also be a potential approach in cases of UCEC. *RIPK2/RIP2* is involved in the nucleosome-binding oligomerization domain-signalling pathway and induces the production of various pro-inflammatory cytokines (IL-1α, IL-6, tumour necrosis factor-α, IL-18), chemokines (monocyte chemoattractant protein-1,

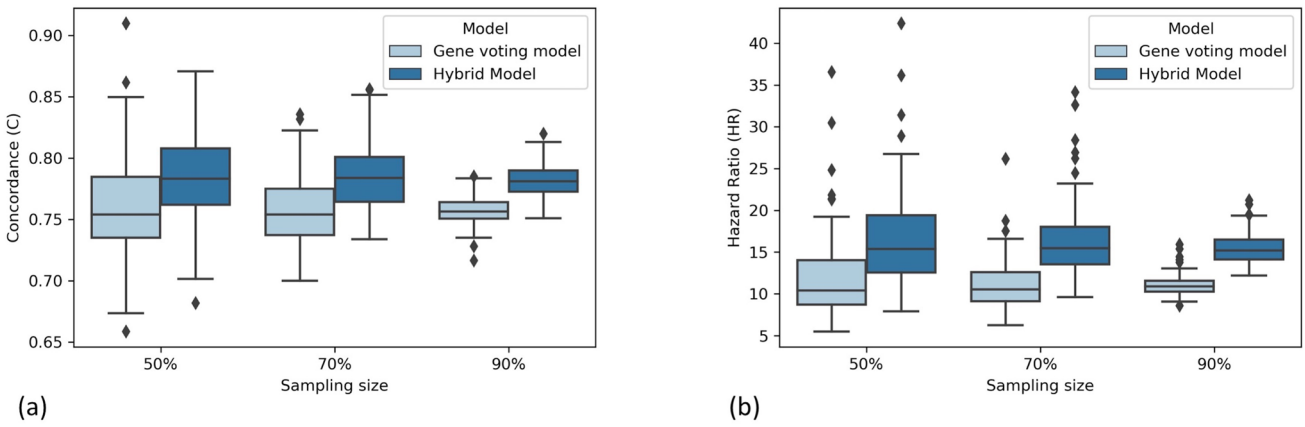

Fig. 8 Predictive validation of voting-based model. **a** Grouped boxplots corresponding to estimated Concordance index (*y*-axis) for 100 iterations of data sampling (*x*-axis). **b** Similarly, estimation of hazard ratio (*y*-axis) for different models using random sampling (*x*-axis)

Gene Voting Model Prediction

True Condition	Low Risk	High Risk
OS > cutoff	True Positive (TP)	False Positive (FP)
OS <= cutoff	False Negative (FN)	True Negative (TN)

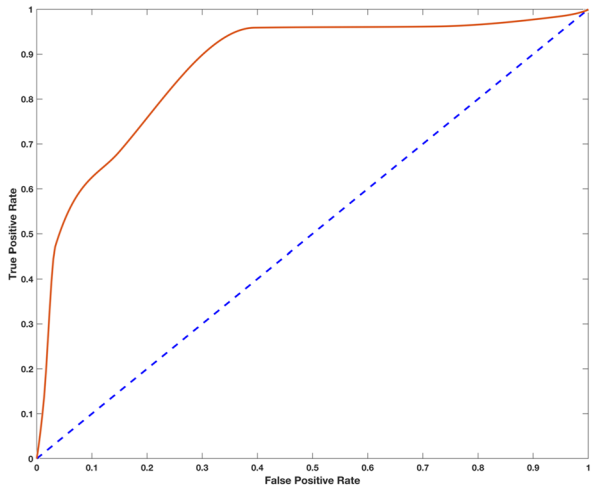

Fig. 9 **a** Terminology used for evaluation of confusion matrix. Initial risk labelling was done using an overall survival (OS) cut-off with patients having OS > cut-off labelled as positive or low risk and vice versa for patients with OS ≤ cut-off. **b** Receiver operating characteristic (ROC) curve for gene voting model area under the ROC curve (AUROC) of 0.86 was obtained.

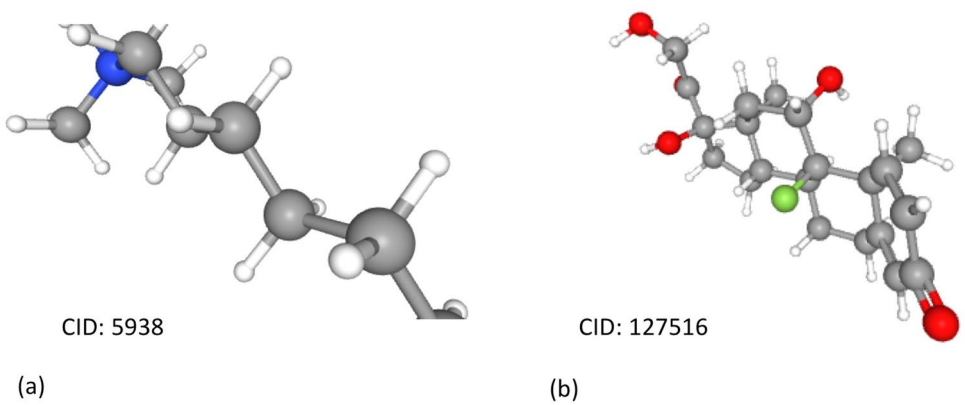

Fig. 10 3D conformer of two most significant small-molecule drugs: **a** Hexamethonium bromide (enrichment = 0.834; $p = 2.6 \times 10^{-4}$) **b** isoflupredone (enrichment = − 0.955; $p = 2.2 \times 10^{-4}$). *CID* Pubchem Compound ID

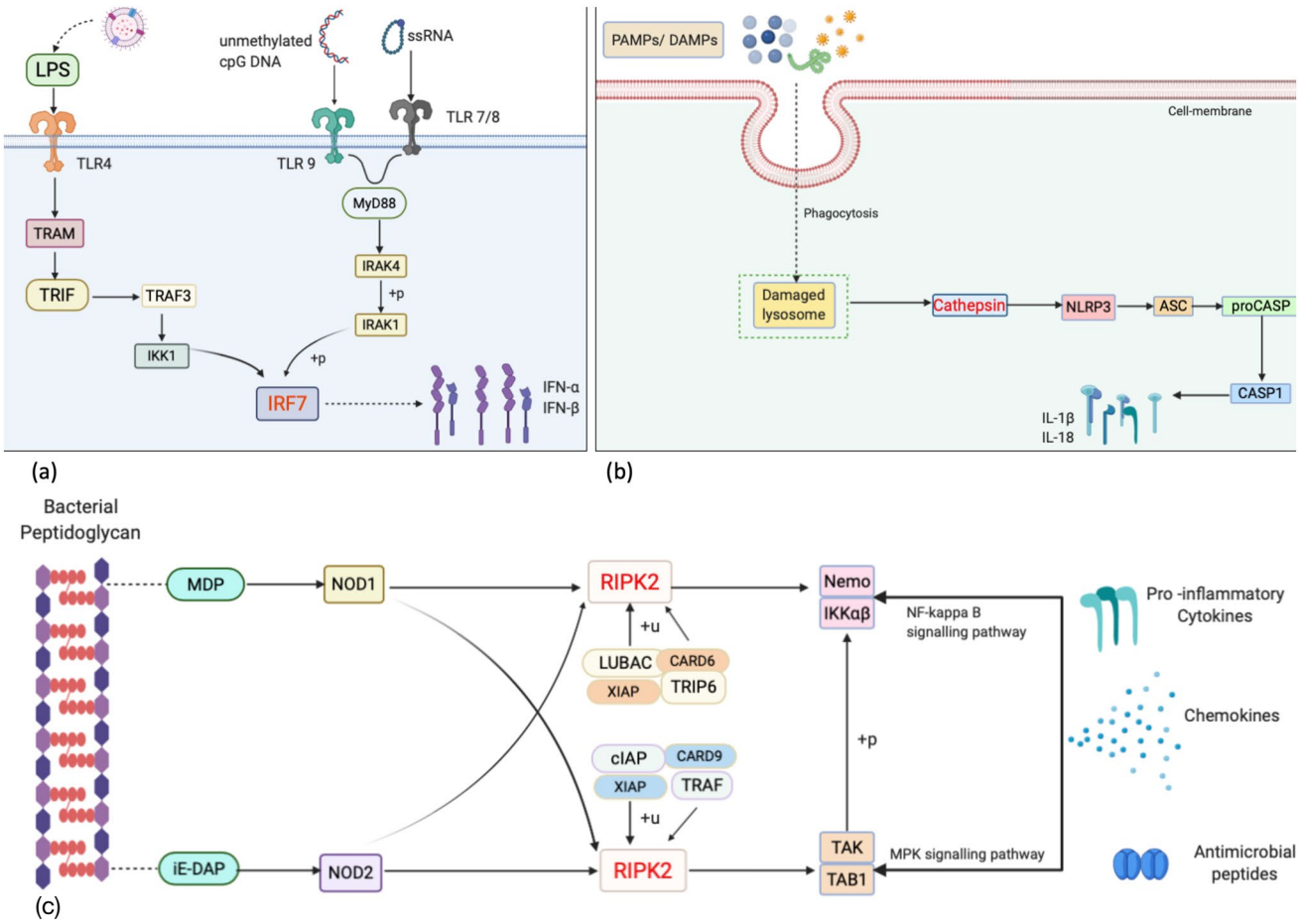

Fig. 11 Schematic illustration of biological mechanism of good prognostic marker (GPM). **a** *IRF-7* **b** *CTSB* (cathepsin), and **c** *RIPK2*. GPM genes are shown in red

chemokine [C-X-C motif] ligand), and antimicrobial peptides (cathelicidin, defensins) via mitogen-activated protein kinase (MAPK) and NF-κβ signalling pathway (Fig. 11c). MESO-DAP1 and mifamurtide (FDA approved) positively regulate *NOD1* and *NOD2*, so can be indirect regulators for *RIPK2* [46]. Bacterial peptidoglycan may also be given as an agonist that can activate the pathway. The BPM genes (*CLEC1B*, *CLEC3A*, *FCN1*, *NLRP10*), which showed poor prognosis in our study, need to be inhibited. Abnormal overexpression of protein kinase B (AKT) has been observed in ovarian, lung, and pancreatic cancers and is associated with increased cancer cell proliferation and survival (Fig. 12). Therefore, targeting AKT could provide an essential approach for cancer prevention and therapy. It has also been seen that the suppression of *CLEC-3A* inhibits cell proliferation and increases chemo-sensitivity through the AKT/mammalian target of rapamycin (mTOR)/hypoxia-inducible factor (HIF)-1α pathway in osteosarcoma [47]. Resveratrol or grape powder (NCT 00256334) is a natural AKT inhibitor in clinical phase I, whereas MK-2206 (NCT 01333475) is a synthetic AKT inhibitor in clinical phase II [48]. These might be used as AKT inhibitors if the trials are successful. These results provided biological plausibility for repositioning cancer drugs for EC pharmacotherapy.

4 Discussion

Although UCEC is known to have a good prognosis if diagnosed early, patients at advanced stage are associated with an abysmal prognosis and high mortality risk. As such, efficient risk assessment strategies are required for clinical decision making and therapeutic intervention. The clinical features, such as tumour grade, cervical involvement, lymph node status, histological subtype, depth of myometrial invasion, and LVSI are significant in risk stratification in UCEC but are not efficient because of their limitations. Thus, aided by the development of

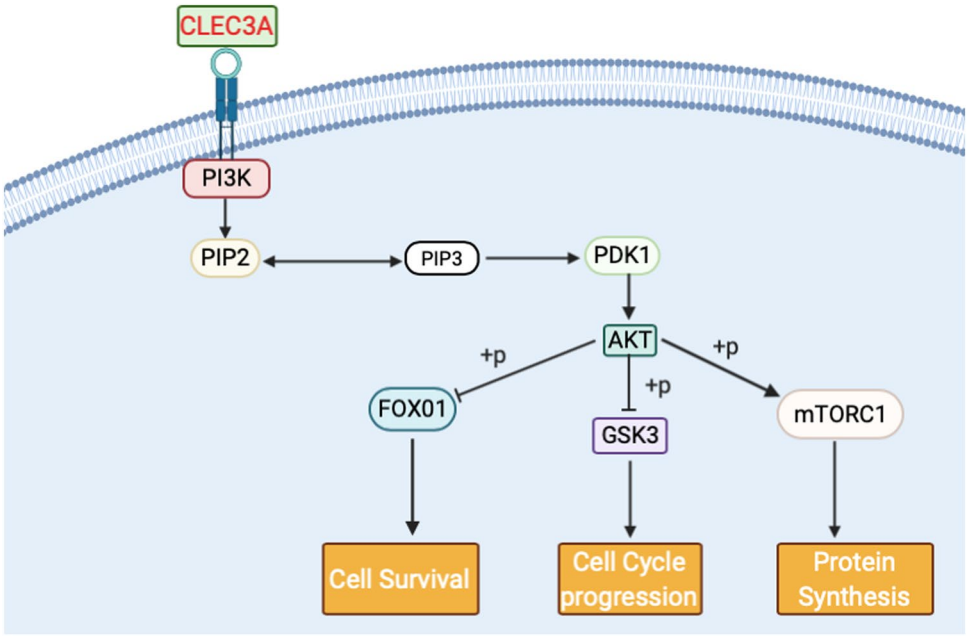

Fig. 12 Schematic illustration of biological mechanism of *CLEC3A*: one of the bad prognostic markers obtained, shown in red

high-throughput sequencing methods and the availability of a massive amount of experimental data, various molecular prognostic markers have been proposed. Earlier studies revealed multiple mechanisms at the molecular level that led to complicated molecular processes that are crucial for cancer progression and development, such as the elucidation of multifaceted signalling processes orchestrated by PRRs. This delineation of the regulatory role of PRRs has boosted therapeutic decision-making strategies in various cancers. PRR agonists are currently used in potential systemic treatments such as chemotherapy, targeted therapy, and immunotherapy in the form of vaccine adjuvants [49] such as AS15, which is a *TLR4/9* agonist and is used as an adjuvant to vaccines dHER2 (truncated form of *HER2*) and lapatinib in breast cancer [50]. Monophosphoryl lipid A, associated with *TLR4*, was shown to act as a potent vaccine adjuvant and to promote type 1 T helper (Th_1)-based immune response in human papillomavirus-induced cervical cancer [51]. Imiquimod associated with *TLR7* was reported to induce apoptosis and stimulate a cell-mediated immune response in basal cell carcinoma [52]. A synthetic derivative of lipid A, OM-174, when associated with *TLR2/4,* is claimed to reduce tumour progression and prolong survival, especially in combination with cyclophosphamide in melanoma [53]. FDA-approved PRR targeted therapies include AS04, a *TLR4* agonist in cervical cancer; imiquimod, a *TLR7/8* agonist, in various skin cancers; and mifamurtide, an *NOD2* agonist in osteosarcoma. Combining PRR-based agonist therapy with immune checkpoint-targeted antibodies (such as anti-cytotoxic T-lymphocyte-associated protein-4 or anti-programmed cell death ligand 1 antibodies) may be the future of cancer therapies [49]. However, the role of PRR signalling genes and their utility in UCEC therapy is still poorly understood.

In this study, we used the messenger RNA (mRNA) expression data obtained from the TCGA-UCEC cohort. First, we estimated the prognostic performance of each PRR gene by employing survival analysis. We used gene co-expression network-based feature selection and a clustering-based approach to find critical PRR genes. Next, we created risk stratification models using the genes obtained from this process. To improve the performance of the model, we identified 15 PRR-related biomarker genes that were associated with UCEC prognosis based on Cox-regression survival analysis: *CLEC1B*, *CLEC3A*, *MRC1*, *IRF7*, *CTSB*, *FCN1*, *RIPK2*, *CLEC3B*, *CLEC12B*, *TLR4*, *NLRP10*, *NLRP9*, *TNIP1*, *SARM1*, and *MAPKAPK2*. A nine-gene (*CLEC1B*, *CLEC3A*, *IRF7*, *CTSB*, *FCN1*, *RIPK2*, *NLRP10*, *NLRP9*, and *SARM1*) voting-based model performed the best and also stratified high-risk clinical groups significantly. Finally, after a comprehensive prognostic comparison with other clinicopathological factors, we developed a hybrid model combining the expression profiles of nine genes with 'clinical stage' to predict high- and low-risk patients with UCEC with high precision. We also predicted candidate biomolecules that can help modulate the gene expressions and also possibly act as drugs in the treatment of UCEC.

The obtained nine main biomarker genes (*CLEC1B*, *CLEC3A*, *IRF7*, *CTSB*, *FCN1*, *RIPK2*, *NLRP10*, *NLRP9*, and *SARM1*) have been shown to exhibit effective regulatory roles in multiple diseases, including cancer. *RIPK2*, *SARM1*, *IRF7*, *CTSB*, and *NLRP9* were linked with good prognosis, whereas *CLEC1B*, *CLEC3A*, *FCN1*, and *NLRP10* were linked with poor prognosis. *CLEC1B*,

or C-type lectin domain family 1 member B, has been shown to be beneficial in atherosclerosis. *CLEC1B* measured plasma concentrations were directly linked with an increased risk of carotid plaque formation. ORs in platelet-derived growth factor receptor-β were 0.79 (95% CI 0.66–0.94; $p = 0.008$) per 1-standard deviation rise [54]. Aggretin α-chain C-terminus is a potential candidate for the treatment of tumour metastasis through *CLEC-2* blockade [55]. *CLEC2* appears to suppress AKT signalling and the invasive activities of gastric cancer cells by blocking expression of phosphoinositide 3-kinase (PI3K) subunits [56]. Overexpression of *CLEC3A* promotes tumour progression and poor prognosis in invasive ductal breast cancer. The knockdown of *CLEC3A* by RNA interference approach efficiently inhibited the proliferation, migration, and invasion of breast cancer cells, which may be mediated by PI3K/AKT signalling pathway. Thus it has an anti-cancer effect and can be a critical therapeutic for breast cancer [57]. *FCN1* can be considered a supplementary biomarker in acute myeloid leukaemia in adults ($p = 0.004$; OR 2.95; 95% CI 1.41–6.16). *FCN1* gene polymorphisms are associated with higher expression of specific mRNA in monocytes and granulocytes as well as higher *FCN1* serum levels [58]. *NLRP10* binds to an apoptosis-associated speck-like protein and can inhibit NF-κB activation and apoptosis as well as caspase-1-mediated IL-1β maturation and thus helps in the regulation of apoptosis and inflammation [59]. SARM mediates intrinsic apoptosis via Bcl-2 family members. It suppresses Bcl extra-large (Bcl-xL) and downregulates extracellular signal-regulated kinase phosphorylation [60]. *SARM1* can act as one of the epigenetic biomarkers in colorectal cancer and helps in the identification of cancer by underexpression and/or CpG methylation [61]. *IRF7* regulates the development of granulocytic myeloid-derived suppressor cells through S100A9 trans-repression in cancer [62]. The functional polymorphism rs12898 in cathepsin B (*CTSB)* may contribute to primary hepatic cancer susceptibility, and the variant A allele could increase the risk of the cancer [63]. *NLRP9* acts as an inflammasome-related molecule, a reliable non-invasive tool for breast cancer diagnosis [64]. The assessment of *NLRP4* and *NLRP9* expression could be beneficial in predicting Bacillus Calmette-Guérin failure and also in decision making for early radical surgery [65].

One study [66] found that, by deleting biomarker genes (and all genes that can be found in the same biological process), new signatures with the same prognostic prediction capabilities but opposing biological meaning could always be found. To understand whether the alternate set of biomarker genes were equally effective in our study, we performed a similar analysis on our initial list of 15 significant biomarker genes. We removed the nine main biomarker genes (involved in our final risk-prediction model) and instead chose the remaining set of six alternate biomarker genes for model development. The best performing model, with HR 4.65 and p value 1.09×10^{-5}, of the various combinations of these six genes comprised five genes. These results indicated that our final model, which consisted of nine main biomarker genes, has a superior performance. To address our next question about the (co-expression-based or biological) correlation between main biomarker genes and alternate biomarker genes, we found that most of the alternate biomarker genes correlated with one of the nine main biomarker genes (Fig. 1 in ESM 3), for example, *MRC1* had a correlation of 0.45 with *FCN1*. This explains why the alternate set of biomarker genes also correlated with survival rather than just the set of main biomarker genes. We also tried finding the biological relationships of these nine and five biomarker genes. We observed that most of these genes have overlapping biological pathways. Figure 2 in ESM 3 shows these results via a heat map showing the intersection of biological pathways among these nine and five other genes from our study. Thus, we concluded that the alternate genes are correlated with the main biomarker genes both in expression and in shared biological mechanisms. Therefore, the model with the main biomarker genes is the optimal choice in our case. In an additional analysis, we clustered the samples into three groups (high risk, mid risk, and low risk) according to their OS time. Thereafter, the expression-based correlation (PCC) amongst gene pairs were evaluated for these three groups. We then took the top ten highly correlated gene pairs from the low-risk group and examined whether the high correlation in these gene pairs was maintained in the mid- and high-risk groups. We observed that PCC values in all gene pairs decreased in the mid-risk group. It further decreased in the high-risk group, as shown in Table 11 in ESM 1. These observations indicate that the correlation between highly correlated gene pairs is disturbed in high-risk patients. In other words, low or no PCC in these gene pairs is potentially associated with poorer survival of patients with UCEC.

Our study incorporated the expression of PRR genes for several UCEC samples and investigated the prognostic significance of each of these in depth. The resultant nine-gene signature was able to classify samples into risk groups and exhibited superior performance when compared with many of the clinicopathological features. These nine biomarker genes can be exploited as therapeutic targets for treatment. This can be achieved through modulation of the expression of these genes to the required low-risk profile. Apart from in silico prediction of small molecules that achieve this goal, we also explored the mechanistic roles of several of these candidate genes and thereby suggested potential agonists for therapeutic repurposing. This work is reasonable, offers a realistic framework for subsequent research, and can serve as a reference for further experimental efforts.

To eliminate the possibility of data bias and establish the robustness of our model, we implemented a predictive validation technique wherein the model's performance was assessed by an iterative sampling of the dataset across various sample sizes. However, to determine their roles and provide a basis for potential clinical application, the biomarkers obtained in this analysis also need to be validated in external cohorts and may be integrated with other available biomarkers. Furthermore, we were only able to use OS to examine patient prognosis because of the lack of data about metastasis and recurrence in the TCGA database, which is another limitation of our study.

5 Conclusion

A risk stratification model was developed that combined nine PRR-related genes that could reliably and efficiently estimate survival outcomes and direct personalised anticancer therapy in patients with UCEC. Our results also suggest that the conjunction of the proposed gene signature with clinical staging provides better prognostication than using staging information alone. Therefore, the findings of our study can also be a potential modification to the existing risk evaluation system in UCEC. Overall, this study finds its importance in the enhancement of our understanding of the oncogenic role of innate immune receptors in UCEC.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s40291-021-00539-1.

Declarations

Funding No sources of funding were used to conduct this study or prepare this manuscript.

Conflicts of Interest Dilraj Kaur, Chakit Arora, and Gajendra Pal Singh Raghava have no conflicts of interest that are directly relevant to the content of this article.

Data availability The data that support the findings of this study are available from the corresponding author on request.

Ethics Approval Not applicable.

Consent Not applicable.

Author Contributions Conceptualisation, DK and GPSR; methodology, DK and CA; formal analysis, DK and CA; investigation, DK and GPSR; code development, DK; visualisation and figs, DK and CA; interpretation of data and results, DK, CA and GPSR; supervision, GPSR; writing and editing, DK, CA and GPSR. All authors have read and agreed to the published version of the manuscript.

References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. CA Cancer J Clin. 2020;70:7–30.
2. Cronin KA, Lake AJ, Scott S, Sherman RL, Noone A-M, Howlader N, et al. Annual report to the nation on the status of cancer, part I: National cancer statistics. Cancer. 2018;124:2785–800.
3. Gottwald L, Pluta P, Piekarski J, Spych M, Hendzel K, Topczewska-Tylinska K, et al. Long-term survival of endometrioid endometrial cancer patients. Arch Med Sci. 2010;6:937–44.
4. Setiawan VW, Yang HP, Pike MC, McCann SE, Yu H, Xiang Y-B, et al. Type I and II endometrial cancers: have they different risk factors? J Clin Oncol. 2013;31:2607–18.
5. Black C, Feng A, Bittinger S, Quinn M, Neesham D, McNally O. Uterine papillary serous carcinoma: a single-institution review of 62 cases. Int J Gynecol Cancer. 2016;26:133–40.
6. de la Rubia EC, Martinez-Garcia E, Dittmar G, Gil-Moreno A, Cabrera S, Colas E. Prognostic biomarkers in endometrial cancer: a systematic review and meta-analysis. J Clin Med. 2020;9:1900.
7. Morice P, Leary A, Creutzberg C, Abu-Rustum N, Darai E. Endometrial cancer. Lancet (London, England). 2016;387:1094–108.
8. Chen P, Yang Y, Zhang Y, Jiang S, Li X, Wan J. Identification of prognostic immune-related genes in the tumor microenvironment of endometrial cancer. Aging (Albany NY). 2020;12:3371–87.
9. Bian J, Xu Y, Wu F, Pan Q, Liu Y. Identification of a five-gene signature for predicting the progression and prognosis of stage I endometrial carcinoma. Oncol Lett. 2020;20:2396–410.
10. Wang Z-H, Zhang Y-Z, Wang Y-S, Ma X-X. Identification of novel cell glycolysis related gene signature predicting survival in patients with endometrial cancer. Cancer Cell Int. 2019;19:296.
11. Han M, Yang HJ, Lin Q. KLHL14, an ovarian and endometrial-specific gene, is over-expressed in ovarian and endometrial cancer. Math Biosci Eng. 2019;17:1702–17.
12. Lata S, Raghava GPS. PRRDB: a comprehensive database of pattern-recognition receptors and their ligands. BMC Genom. 2008;9:180.
13. Kaur D, Patiyal S, Sharma N, Usmani SS, Raghava GPS. PRRDB 2.0: a comprehensive database of pattern-recognition receptors and their ligands. Database. 2019;2019:baz076.
14. Goutagny N, Estornes Y, Hasan U, Lebecque S, Caux C. Targeting pattern recognition receptors in cancer immunotherapy. Target Oncol. 2012;7:29–54.
15. Ikebe M, Kitaura Y, Nakamura M, Tanaka H, Yamasaki A, Nagai S, et al. Lipopolysaccharide (LPS) increases the invasive ability of pancreatic cancer cells through the TLR4/MyD88 signaling pathway. J Surg Oncol. 2009;100:725–31.
16. Alvero AB, Chen R, Fu H-H, Montagna M, Schwartz PE, Rutherford T, et al. Molecular phenotyping of human ovarian cancer stem cells unravels the mechanisms for repair and chemoresistance. Cell Cycle. 2009;8:158–66.
17. Sheyhidin I, Nabi G, Hasim A, Zhang R-P, Ainiwaer J, Ma H, et al. Overexpression of TLR3, TLR4, TLR7 and TLR9 in esophageal squamous cell carcinoma. World J Gastroenterol. 2011;17:3745–51.
18. Gu J, Liu Y, Xie B, Ye P, Huang J, Lu Z. Roles of toll-like receptors: from inflammation to lung cancer progression. Biomed Rep. 2018;8:126–32.
19. Grimmig T, Matthes N, Hoeland K, Tripathi S, Chandraker A, Grimm M, et al. TLR7 and TLR8 expression increases tumor cell proliferation and promotes chemoresistance in human pancreatic cancer. Int J Oncol. 2015;47:857–66.

20. Lanki M, Seppanen H, Mustonen H, Hagstrom J, Haglund C. Toll-like receptor 1 predicts favorable prognosis in pancreatic cancer. PLoS ONE. 2019;14:e0219245.
21. Ronkainen H, Hirvikoski P, Kauppila S, Vuopala KS, Paavonen TK, Selander KS, et al. Absent Toll-like receptor-9 expression predicts poor prognosis in renal cell carcinoma. J Exp Clin Cancer Res. 2011;30:84.
22. Wang C, Cao S, Yan Y, Ying Q, Jiang T, Xu K, et al. TLR9 expression in glioma tissues correlated to glioma progression and the prognosis of GBM patients. BMC Cancer. 2010;10:415.
23. Tuomela J, Sandholm J, Karihtala P, Ilvesaro J, Vuopala KS, Kauppila JH, et al. Low TLR9 expression defines an aggressive subtype of triple-negative breast cancer. Breast Cancer Res Treat. 2012;135:481–93.
24. Li Z, Block MS, Vierkant RA, Fogarty ZC, Winham SJ, Visscher DW, et al. The inflammatory microenvironment in epithelial ovarian cancer: a role for TLR4 and MyD88 and related proteins. Tumour Biol. 2016;37:13279–86.
25. Ashton KA, Proietto A, Otton G, Symonds I, McEvoy M, Attia J, et al. Toll-like receptor (TLR) and nucleosome-binding oligomerization domain (NOD) gene polymorphisms and endometrial cancer risk. BMC Cancer. 2010;10:382.
26. Kelly MG, Alvero AB, Chen R, Silasi D-A, Abrahams VM, Chan S, et al. TLR-4 signaling promotes tumor growth and paclitaxel chemoresistance in ovarian cancer. Cancer Res. 2006;66:3859–68.
27. Allhorn S, Boing C, Koch AA, Kimmig R, Gashaw I. TLR3 and TLR4 expression in healthy and diseased human endometrium. Reprod Biol Endocrinol. 2008;6:40.
28. van der Net JB, Janssens AC, Defesche JC, Kastelein JJ, Sijbrands EJ, Steyerberg EW. Usefulness of genetic polymorphisms and conventional risk factors to predict coronary heart disease in patients with familial hypercholesterolemia. Am J Cardiol. 2009;103(3):375–80.
29. Kaur D, Arora C, Raghava GPS. A hybrid model for predicting pattern recognition receptors using evolutionary information. Front Immunol. 2020;11:71.
30. Li P, Ren H, Zhang Y, Zhou Z. Fifteen-gene expression based model predicts the survival of clear cell renal cell carcinoma. Medicine (Baltimore). 2018;97(33):e11839.
31. Wang Y, Ren F, Chen P, Liu S, Song Z, Ma X. Identification of a six-gene signature with prognostic value for patients with endometrial carcinoma. Cancer Med. 2018;7(11):5632–42.
32. Abbott KL, Nyre ET, Abrahante J, Ho Y-Y, Isaksson Vogel R, Starr TK. The candidate cancer gene database: a database of cancer driver genes from forward genetic screens in mice. Nucleic Acids Res. 2015;43:D844–8.
33. Zhou H, Zhang C, Li H, Chen L, Cheng X. A novel risk score system of immune genes associated with prognosis in endometrial cancer. Cancer Cell Int. 2020;20:240.
34. Zhao N, Guo M, Wang K, Zhang C, Liu X. Identification of pan-cancer prognostic biomarkers through integration of multi-omics data. Front Bioeng Biotechnol. 2020;8:268.
35. Musa A, Ghoraie LS, Zhang S-D, Glazko G, Yli-Harja O, Dehmer M, et al. A review of connectivity map and computational approaches in pharmacogenomics. Brief Bioinform. 2018;19:506–23.
36. Shen Y, Dong S, Liu J, Zhang L, Zhang J, Zhou H, et al. Identification of potential biomarkers for thyroid cancer using bioinformatics strategy: a study based on GEO datasets. Biomed Res Int. 2020;2020:9710421.
37. Mackey WA, Shaw GB. Oral hexamethonium bromide in essential hypertension. Br Med J. 1951;2:259–65.
38. Wang Y, Bryant SH, Cheng T, Wang J, Gindulyte A, Shoemaker BA, et al. PubChem BioAssay: 2017 update. Nucleic Acids Res. 2017;45:D955–63.
39. Abdolvahab MH, Darvishi B, Zarei M, Majidzadeh-A K, Farahmand L. Interferons: role in cancer therapy. Immunotherapy. 2020;12:833–55.
40. Gogas H, Ioannovich J, Dafni U, Stavropoulou-Giokas C, Frangia K, Tsoutsos D, et al. Prognostic significance of autoimmunity during treatment of melanoma with interferon. N Engl J Med. 2006;354:709–18.
41. Qin XQ, Tao N, Dergay A, Moy P, Fawell S, Davis A, et al. Interferon-beta gene therapy inhibits tumor formation and causes regression of established tumors in immune-deficient mice. Proc Natl Acad Sci USA. 1998;95:14411–6.
42. Rohatiner AZS, Gregory WM, Peterson B, Borden E, Solal-Celigny P, Hagenbeek A, et al. Meta-analysis to evaluate the role of interferon in follicular lymphoma. J Clin Oncol. 2005;23:2215–23.
43. Locatelli MC, Facendola G, Pizzocaro G, Piva L, Pegoraro C, Pallavicini EB, et al. Subcutaneous administration of interleukin-2 and interferon-alpha 2b in advanced renal cell carcinoma: long-term results. Cancer Detect Prev. 1999;23:172–6.
44. Solinhac B. Augmentation materials in surgery. Can ridge height be increased in edentulous patients? Rev Odontostomatol. 1985;14:247–50.
45. Novak N, Yu C-F, Bieber T, Allam J-P. Toll-like receptor 7 agonists and skin. Drug News Perspect. 2008;21:158–65.
46. Hasuo H, Akasu T. Activation of inhibitory pathways suppresses the induction of long-term potentiation in neurons of the rat lateral septal nucleus. Neuroscience. 2001;105:343–52.
47. Ren C, Pan R, Hou L, Wu H, Sun J, Zhang W, et al. Suppression of CLEC3A inhibits osteosarcoma cell proliferation and promotes their chemosensitivity through the AKT1/mTOR/HIF1alpha signaling pathway. Mol Med Rep. 2020;21:1739–48.
48. Song M, Bode AM, Dong Z, Lee M-H. AKT as a therapeutic target for cancer. Cancer Res. 2019;79:1019–31.
49. Bai L, Li W, Zheng W, Xu D, Chen N, Cui J. Promising targets based on pattern recognition receptors for cancer immunotherapy. Pharmacol Res. 2020;159:105017.
50. Milani A, Sangiolo D, Montemurro F, Aglietta M, Valabrega G. Active immunotherapy in HER2 overexpressing breast cancer: current status and future perspectives. Ann Oncol. 2013;24:1740–8.
51. Gregg KA, Harberts E, Gardner FM, Pelletier MR, Cayatte C, Yu L, et al. Rationally designed TLR4 ligands for vaccine adjuvant discovery. MBio. 2017;8:e00492-e517.
52. Bubna AK. Imiquimod—its role in the treatment of cutaneous malignancies. Indian J Pharmacol. 2015;47:354–9.
53. D'Agostini C, Pica F, Febbraro G, Grelli S, Chiavaroli C, Garaci E. Antitumour effect of OM-174 and cyclophosphamide on murine B16 melanoma in different experimental conditions. Int Immunopharmacol. 2005;5:1205–12.
54. Mosley JD, Benson MD, Smith JG, Melander O, Ngo D, Shaffer CM, et al. Probing the virtual proteome to identify novel disease biomarkers. Circulation. 2018;138:2469–81.
55. Chang C-H, Chung C-H, Hsu C-C, Peng H-C, Huang T-F. Inhibitory effects of polypeptides derived from a snake venom C-type lectin, aggretin, on tumor cell-induced platelet aggregation. J Thromb Haemost. 2014;12:540–9.
56. Wang L, Yin J, Wang X, Shao M, Duan F, Wu W, et al. C-type lectin-like receptor 2 suppresses AKT signaling and invasive activities of gastric cancer cells by blocking expression of phosphoinositide 3-kinase subunits. Gastroenterology. 2016;150:1183-1195.e16.
57. Ni J, Peng Y, Yang F-L, Xi X, Huang X-W, He C. Overexpression of CLEC3A promotes tumor progression and poor prognosis in breast invasive ductal cancer. Onco Targets Ther. 2018;11:3303–12.

58. Sokolowska A, Swierzko AS, Gajek G, Golos A, Michalski M, Nowicki M, et al. Associations of ficolins and mannose-binding lectin with acute myeloid leukaemia in adults. Sci Rep. 2020;10:10561.
59. Zambetti LP, Laudisi F, Licandro G, Ricciardi-Castagnoli P, Mortellaro A. The rhapsody of NLRPs: master players of inflammation and a lot more. Immunol Res. 2012;53:78–90.
60. Panneerselvam P, Singh LP, Selvarajan V, Chng WJ, Ng SB, Tan NS, et al. T-cell death following immune activation is mediated by mitochondria—localized SARM. Cell Death Differ. 2013;20:478–89.
61. Quyun C, Ye Z, Lin S-C, Lin B. Recent patents and advances in genomic biomarker discovery for colorectal cancers. Recent Pat DNA Gene Seq. 2010;4:86–93.
62. Yang Q, Li X, Chen H, Cao Y, Xiao Q, He Y, et al. IRF7 regulates the development of granulocytic myeloid-derived suppressor cells through S100A9 transrepression in cancer. Oncogene. 2017;36:2969–80.
63. Cui M, Chen Q, He C, Wang N, Yu Y, Sun Z, et al. A single nucleotide polymorphism CTSB rs12898 is associated with primary hepatic cancer in a Chinese population. Int J Clin Exp Pathol. 2019;12:3063–9.
64. Mearini E, Poli G, Cochetti G, Boni A, Egidi MG, Brancorsini S. Expression of urinary miRNAs targeting NLRs inflammasomes in bladder cancer. Onco Targets Ther. 2017;10:2665–73.
65. Poli G, Cochetti G, Boni A, Egidi MG, Brancorsini S, Mearini E. Characterization of inflammasome-related genes in urine sediments of patients receiving intravesical BCG therapy. Urol Oncol. 2017;35(12):674.e19-674.e24.
66. Manjang K, Tripathi S, Yli-Harja O, Dehmer M, Glazko G, Emmert-Streib F. Prognostic gene expression signatures of breast cancer are lacking a sensible biological meaning. Sci Rep. 2021;11:156.