



OPEN ACCESS

EDITED BY

Ashok Kumar,
All India Institute of Medical Sciences,
Bhopal, India

REVIEWED BY

Yang Chen,
Chinese Academy of Sciences (CAS),
China
Yu'e Liu,
Tongji University, China

*CORRESPONDENCE

Jin Li,
✉ 65887512@qq.com
Wenzhong Zhao,
✉ zhaowzphd@163.com

RECEIVED 23 March 2023

ACCEPTED 01 June 2023

PUBLISHED 09 June 2023

CITATION

Wu R, Wu C, Zhu B, Li J and Zhao W
(2023), Screening and validation of
potential markers associated with uterine
corpus endometrial carcinoma and
polycystic ovary syndrome based on
bioinformatics methods.
Front. Mol. Biosci. 10:1192313.
doi: 10.3389/fmolb.2023.1192313

COPYRIGHT

© 2023 Wu, Wu, Zhu, Li and Zhao. This is
an open-access article distributed under
the terms of the [Creative Commons
Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use,
distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication
in this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

Screening and validation of potential markers associated with uterine corpus endometrial carcinoma and polycystic ovary syndrome based on bioinformatics methods

Ruishan Wu¹, Cailin Wu², Bingming Zhu³, Jin Li^{4*} and Wenzhong Zhao^{1*}

¹NHC Key Laboratory of Male Reproduction and Genetics, Guangdong Provincial Reproductive Science Institute (Guangdong Provincial Fertility Hospital), Guangzhou, China, ²Department of Gynecology, The University of HongKong–Shenzhen Hospital, Shenzhen, China, ³Department of Clinical Laboratory, The First Affiliated Hospital of Jinan University, Guangzhou, China, ⁴Department of Pain Management, The First Affiliated Hospital of Jinan University, Guangzhou, China

Background: Endometrial cancer (UCEC) is a commonly occurring tumor in females, and polycystic ovary syndrome (PCOS) is closely related to UCEC, but the molecular mechanisms remain unclear. This article aims to explore potential molecular mechanisms in UCEC and PCOS, as well as identify prognostic genes for UCEC.

Methods: Bioinformatics methods were employed to screen for DEGs in UCEC and PCOS. The shared DEGs were analyzed by constructing a protein-protein interaction (PPI) network using the String database and Cytoscape software. The enrichment analysis was performed using Metascape. The shared DEGs associated with the prognosis of UCEC were identified through univariate and lasso Cox regression methods. A multivariate Cox regression model was constructed and internally validated. The expression and test efficiency of the key prognostic genes were verified using external datasets for UCEC and PCOS. Furthermore, the Gepia database was utilized to analyze the expression of key prognostic genes and their correlation with the disease-free survival (RFS) of UCEC. Tumor mutation burden (TMB), immune infiltration, and the correlation of immune cells were assessed for the prognostic genes of UCEC.

Results: There were 151 shared DEGs identified between UCEC and PCOS through bioinformatics screening. These shared DEGs were primarily enriched in leukocyte activation. Following model construction and verification, nine genes were determined to be prognostic for UCEC from the shared DEGs. Among them, TSPYL5, KCNJ15, RTN1, HMOX1, DCAF12L1, VNN2, and ANXA1 were confirmed as prognostic genes in UCEC through external validation. Additionally, RTN1 was identified as a key gene in both UCEC and PCOS. Gepia analysis revealed that higher expression of RTN1 was associated with RFS in UCEC. Immune infiltration analysis of the shared DEGs demonstrated significant differences in the expression of various immune cells between UCEC high and low TMB groups. The seven key prognostic genes in UCEC exhibited regulatory relationships with immune cells.

Conclusion: This study identified TSPYL5, KCNJ15, RTN1, HMOX1, DCAF12L1, VNN2, and ANXA1 as the key prognostic DEGs of UCEC. These genes are associated with UCEC survival, TMB, immune cell infiltration, and immune cell regulation. Among them, RTN1 may serve as a potential biomarker for both UCEC and PCOS.

KEYWORDS

endometrial carcinoma (UCEC), polycystic ovary syndrome (PCOS), prognostic marker, tumor mutation burden, RTN1

1 Introduction

Endometrial cancer (UCEC) is a common malignant tumor in the female reproductive system, specifically in the endometrium. It is more prevalent in developed countries, particularly among middle-aged and menopausal women (Crosbie et al., 2022). The incidence of endometrial cancer is 5.9% (Banz-Jansen et al., 2022) and its mortality rate has been increasing by an average of 1.9% annually (Makker et al., 2021). Obesity is a recognized risk factor for the most common type of UCEC (Rahib et al., 2014; Lauby-Secretan et al., 2016). Polycystic ovary syndrome (PCOS) is a multifactorial, multigenic, inflammatory autoimmune disorder that commonly affects reproductive-age women. It is characterized by hyperandrogenism, insulin resistance, ovulation disorders, metabolic disorders, and other features (Liao et al., 2021). PCOS often leads to health problems such as obesity, infertility, metabolic disorders, cardiovascular diseases, and cancers (Patel, 2018; Liao et al., 2021). Obesity and complex metabolic diseases, including PCOS and diabetes, are considered risk factors for UCEC (Lauby-Secretan et al., 2016; Patel, 2018; Liao et al., 2021).

UCEC is pathologically staged into four grades by the International Federation of Gynecology and Obstetrics. A higher grade indicates greater cancer cell proliferation and more challenging treatment. The late-stage survival rate of UCEC is only 17%, highlighting the importance of early diagnosis and treatment in improving survival and prognosis for UCEC patients (Zhang M. et al., 2022). UCEC is commonly treated with surgical resection, hormone therapy, radiotherapy, and chemotherapy. However, in recent years, immunotherapy has emerged as a promising new treatment modality for cancer (Riley et al., 2019; Atsavaprane et al., 2021; El-Mayta et al., 2021). By leveraging checkpoint inhibitors and adoptive cell therapy to manipulate the immune system and enhance its recognition and destruction of cancer cells, cancer immunotherapies offer the potential for durable therapeutic responses in a range of solid tumors and hematologic malignancies (Kennedy and Salama, 2020). Immunotherapy is now being applied to various cancers, including UCEC, providing hope for patients (Makker et al., 2019; Liu et al., 2021b).

High-throughput sequencing can simultaneously detect a large amount of genomic information, with fast and accurate detection speed. This capability greatly enhances people's ability to assess the risk of complex diseases and the accuracy of targeted drug treatment (Rodriguez and Miller, 2014; Rego and Snyder, 2019). In recent years, high-throughput sequencing has become widely utilized in the medical field. This method enables a profound understanding of disease pathogenesis from a genetic perspective. (Rego and Snyder,

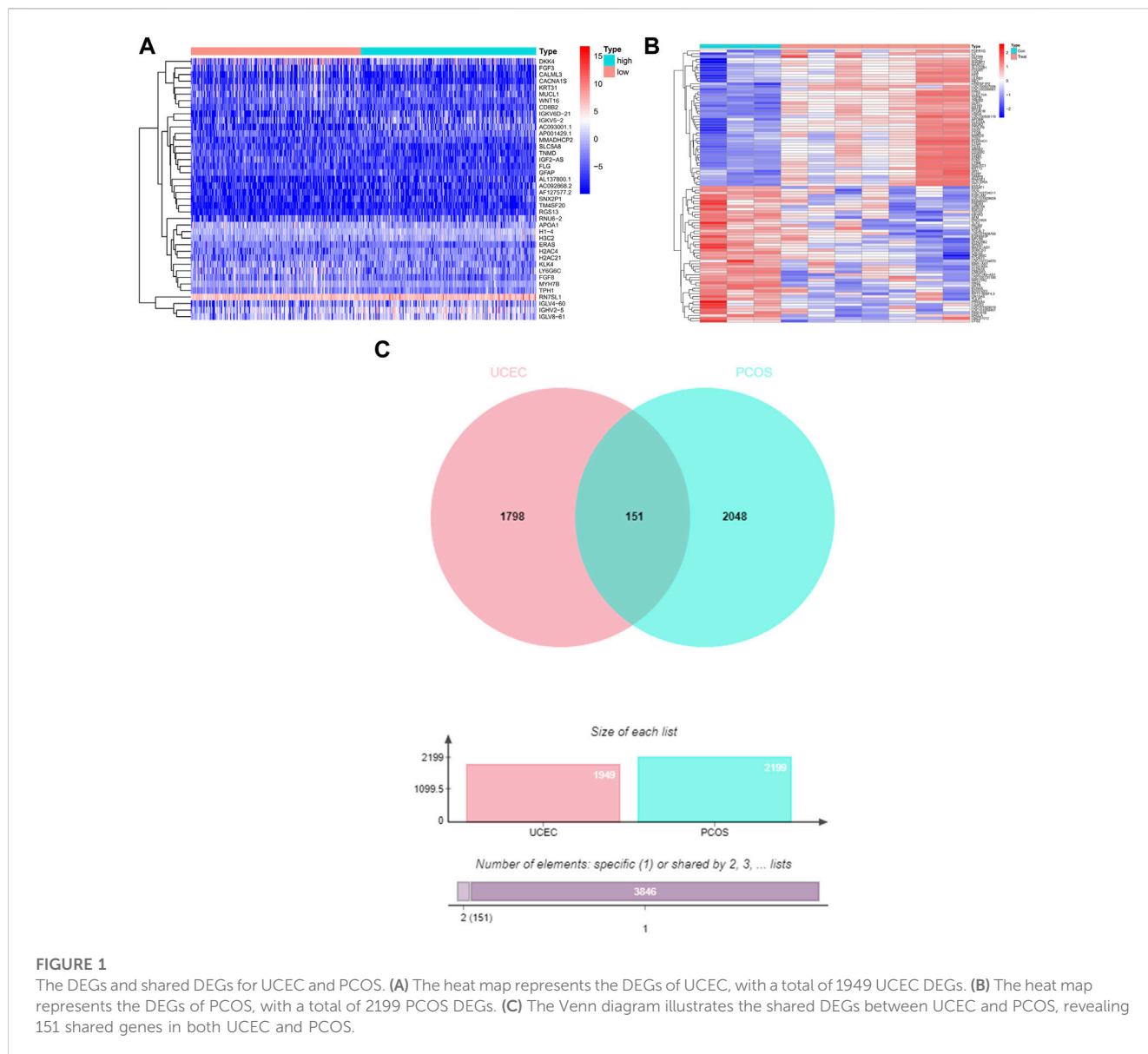
2019). With the advancement and application of high-throughput sequencing technology, the diagnosis and treatment of UCEC have shifted from histology to genomics. Genomic evidence, along with clinicopathological criteria, provides crucial information for UCEC treatment and enables the delivery of improved quality and personalized services to patients (Mitric and Bernardini, 2022). Bioinformatics, combined with high-throughput sequencing, can efficiently screen and verify key disease-related genes. Moreover, it can predict disease risk.

While PCOS has been linked to an increased risk of UCEC (Lauby-Secretan et al., 2016; Patel, 2018; Liao et al., 2021), meta-analyses have shown a significant association between PCOS and the pathogenesis of UCEC (Barry et al., 2014; Li Z. et al., 2022). However, the molecular mechanisms underlying this link remain unclear. Understanding the pathophysiological connection between UCEC and PCOS is crucial for developing effective clinical treatment strategies. To address this, we employed bioinformatics methods to screen for shared DEGs between UCEC and PCOS. We visualized the protein-protein interaction (PPI) network of these genes using the String database and Cytoscape software. Next, we utilized lasso Cox regression, and univariate and multivariate Cox regression methods to identify key shared differentially expressed genes (DEGs) associated with the prognosis of UCEC. We then constructed and validated the model. Using the Gepia database, we analyzed the expression levels and RFS of key shared DEGs across different pathological stages of UCEC. Finally, we conducted tumor mutation burden (TMB) and immune infiltration analysis on both the shared and key prognostic genes for UCEC and PCOS.

2 Materials and methods

2.1 Data download

The data for UCEC were obtained by searching the keywords "transcriptome profiling, Gene Expression Quantification, TCGA-UCEC" from The Cancer Genome Atlas (TCGA) database (<https://portal.gdc.cancer.gov/>). Keywords such as "polycystic ovary syndrome, PCOS" were searched in the Gene Expression Omnibus (GEO) database to retrieve the gene expression profile of PCOS. Additionally, keywords such as "Endometrial cancer, EC, UCEC" were searched in the GEO database to obtain the gene expression profile of UCEC. The inclusion criteria for the PCOS and UCEC datasets were as follows: 1) Gene expression profiles must include both a case group and a control group. 2) Datasets should have available raw data or analyzable data.



2.2 Screening for DEGs

The UCEC data matrix file was downloaded from the TCGA database, the PCOS data matrix file was downloaded from the GEO database, and the “limma” R package, a microarray linear model, was utilized to correct the data and analyze the fold change (FC). The log₂ FC of each gene was used to determine its rank in the final gene list. Genes with a $p < 0.05$ and $|\log_2 FC| \geq 0.5$ were considered significant DEGs and were subjected to further analysis.

2.3 Identifying shared DEGs for UCEC and PCOS

A Venn diagram was constructed using the online website (<http://www.ehbio.com/test/venn/#/>) to identify the intersection

of DEGs between UCEC and PCOS, representing the shared DEGs for both conditions.

2.4 Construction and enrichment analysis of shared DEGs’ PPI network

“String” is an online network tool that can predict protein association between proteins and draw a PPI network (von Mering et al., 2003; Szklarczyk et al., 2017; Szklarczyk et al., 2023). In this study, we utilized String to analyze the protein interaction network of shared DEGs between UCEC and PCOS. The String using method is as follows: Log in to the String database (<https://string-db.org/>), click “Multiple Proteins”, enter shared DEGs into “List of Names”, select “Homo sapiens” for organisms, and click “search” to perform protein mutual assistance on shared

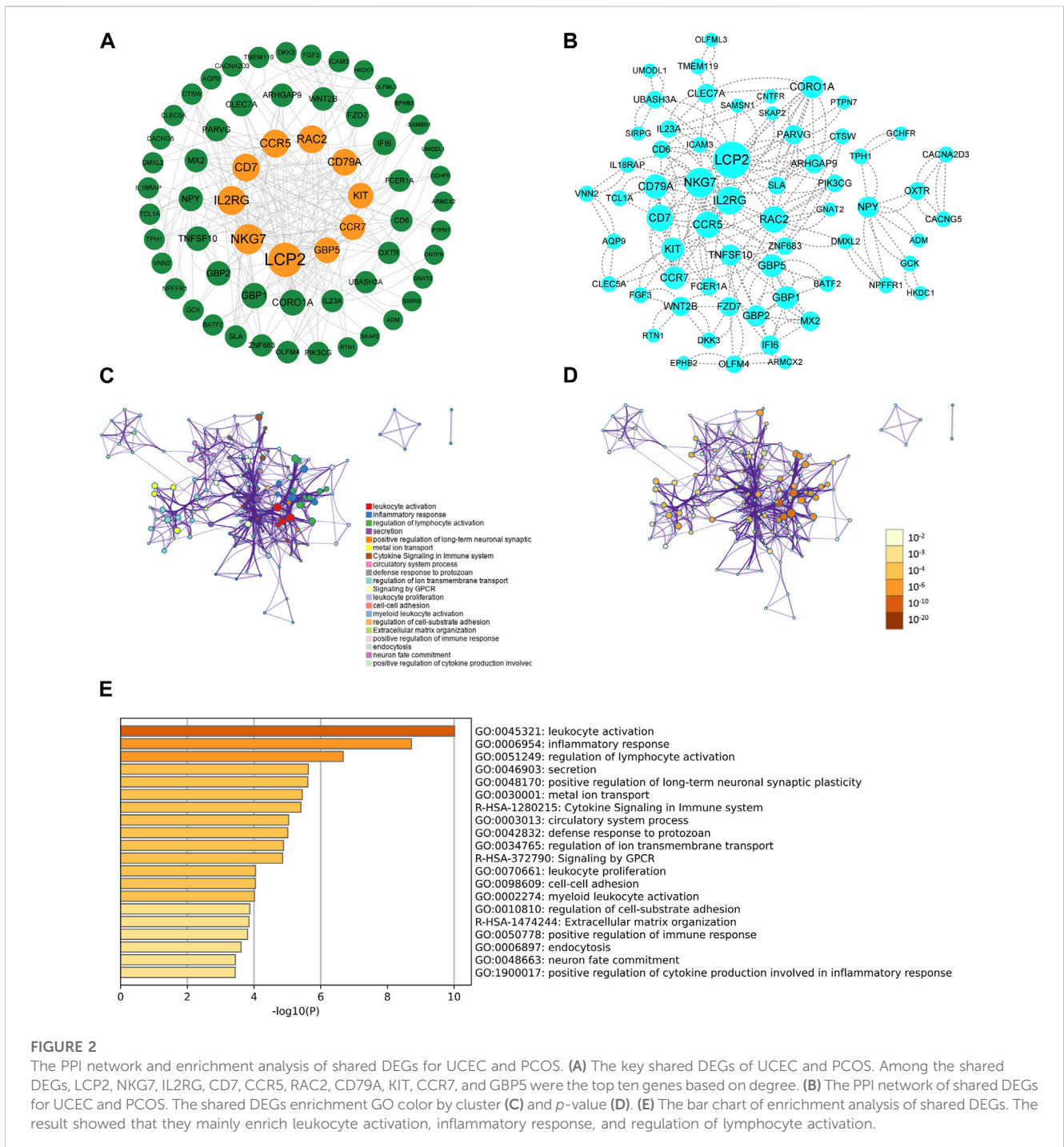


FIGURE 2

The PPI network and enrichment analysis of shared DEGs for UCEC and PCOS. (A) The key shared DEGs of UCEC and PCOS. Among the shared DEGs, LCP2, NKG7, IL2RG, CD7, CCR5, RAC2, CD79A, KIT, CCR7, and GBP5 were the top ten genes based on degree. (B) The PPI network of shared DEGs for UCEC and PCOS. The shared DEGs enrichment GO color by cluster (C) and *p*-value (D). (E) The bar chart of enrichment analysis of shared DEGs. The result showed that they mainly enrich leukocyte activation, inflammatory response, and regulation of lymphocyte activation.

DEGs network analysis, when greater than 0.4 indicates statistical significance.

Cytoscape software, when combined with extensive databases of PPI, protein-DNA, and genetic interactions, is a powerful tool (Shannon et al., 2003). In this study, the String database and Cytoscape software were employed to map and visualize the PPI networks of shared DEGs between UCEC and PCOS. The CytoNCA plug-in in Cytoscape software (version 3.9.1) was utilized to perform topological analysis and betweenness centrality (BC) screening of the resulting

protein interaction network data (Tang et al., 2015). Subsequently, the shared DEGs were ranked based on their degree values. A PPI network map representing the shared DEGs between UCEC and PCOS was generated, and the top 10 shared DEGs with the highest degree values were identified and visualized.

Metascape (<https://metascape.org/gp/index.html#/main/step1>) is an online platform that provides comprehensive annotation and analysis resources for gene lists (Zhou et al., 2019). In this study, we utilized Metascape to perform gene ontology (GO) and

TABLE 1 The top ten degrees of shared DEGs in PCOS and UCEC.

Name	Betweenness	Closeness	Degree	Shared name
LCP2	1247.608	0.371951	28	LCP2
NKG7	438.321	0.308081	20	NKG7
IL2RG	843.6401	0.365269	18	IL2RG
CD7	157.3024	0.309645	16	CD7
CCR5	1065.951	0.335165	16	CCR5
RAC2	1717.966	0.363095	16	RAC2
CD79A	126.3662	0.306533	14	CD79A
KIT	168.2864	0.283721	12	KIT
CCR7	59.69935	0.291866	12	CCR7
GBP5	333.5	0.262931	12	GBP5

Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses on the shared DEGs between UCEC and PCOS.

2.5 Establishment of the UCEC prognostic gene model

The UCEC prognostic model was developed using shared DEGs data from UCEC and PCOS. Univariate Cox analysis was employed to identify UCEC prognostic genes. Lasso analysis and cross-validation were used to address the issue of overfitting by eliminating highly correlated genes. Subsequently, multivariate Cox regression was performed to further reduce dimensionality. Genes associated with UCEC survival in multivariate Cox regression were extracted as prognostic genes. These prognostic genes were utilized to construct the model and generate forest plots.

2.6 Internal validation of the prognostic model on the TCGA database

The TCGA-UCEC data were divided into training (50%) and testing datasets (50%) for model internal validation. The obtained model was then used to predict the risk of training and testing samples. Based on the median value of the training dataset, patients in both datasets were divided into high-risk and low-risk groups. The risk value was calculated using the formula: Risk value = expression level of gene 1 * coef1 + expression level of gene 2 * coef2 + ... expression level of gene n * coefn. To compare the difference in overall survival between the high-risk and low-risk groups, Kaplan-Meier (KM) analysis from the “survival” package was performed on the training and testing datasets. Additionally, the predictive performance of the model was evaluated using receiver operating characteristic (ROC) curves generated for 1, 3, and 5 years using the “survivalROC” package. These ROC curves were created for both the training and testing datasets. Risk values for the training and testing datasets were obtained and used to generate risk score distribution maps, survival distribution maps, and expression heat maps to visualize the results.

2.7 External validation of prognostic genes, and the establishment of key genes

External validation datasets for UCEC and PCOS prognostic genes were obtained from the GEO database, meeting the selection criteria. The expressions of key shared DEGs from UCEC and PCOS were verified using the external validation datasets. UCEC prognostic key genes were selected based on significant differences between normal controls and UCEC in the external validation dataset. Similarly, PCOS key genes were chosen by identifying the key shared DEGs that displayed significant differences between normal controls and PCOS in the external validation dataset. The intersection of externally validated key genes for both UCEC and PCOS represents potential markers for both diseases. The predictive performance of the genes was assessed using ROC curves generated with the “qROC” package. A ROC curve greater than 0.6 was considered indicative of good predictive performance for the gene.

2.8 Expression and survival analysis by Gepia

The Gepia database (<http://gepia.cancer-pku.cn/>) is a valuable resource for data mining and gaining insights into gene function (Tang et al., 2017). In this study, we utilized the “survival” module of the Gepia database to explore the association between the expression of key prognostic genes and the RFS in UCEC.

2.9 Immune infiltration analysis of shared DEGs

Cibersort immune cell infiltration analysis was conducted on the shared DEGs using the “limma” R package. A bar plot distribution of the shared DEGs was generated for 22 immune cell subsets. The shared genes were then categorized into high and low TMB groups in UCEC, and the differences in immune cell infiltrate between these groups were compared.

2.10 Immune correlates of key prognostic genes

To examine the correlation between key prognostic genes in UCEC and immune cells, we utilized the R packages “reshape2”, “ggpubr”, and “ggExtra”. Subsequently, we employed R software to create a lollipop diagram to visualize the correlation between the key prognostic genes and immune cells.

3 Results

3.1 Identification of DEGs

The TCGA-UCEC database consists of 35 cases of normal endometrium and 554 cases of UCEC. Through screening, a total of 1949 UCEC DEGs were identified, including 958 upregulated differential genes and 991 downregulated differential genes. The

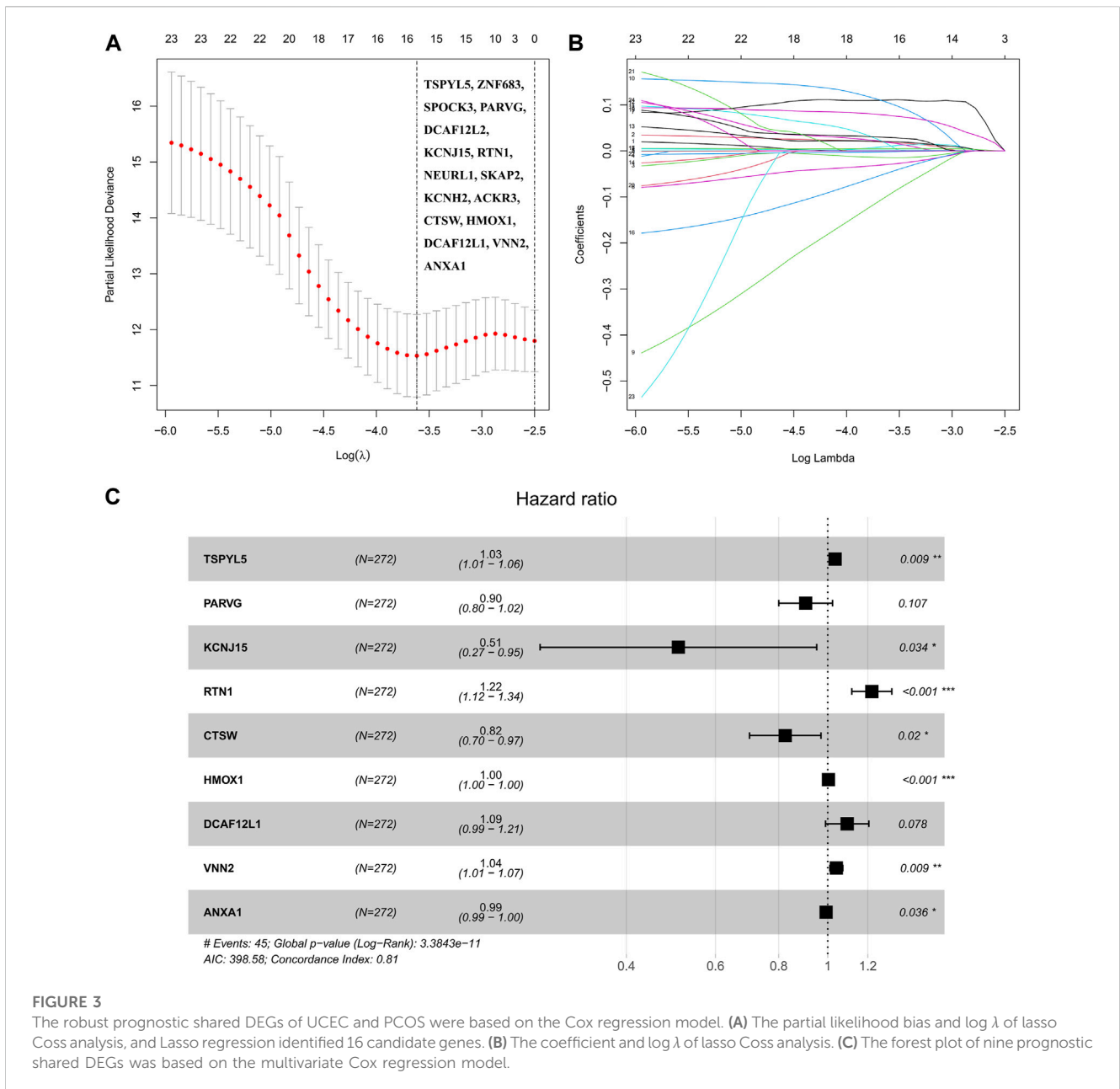


FIGURE 3 The robust prognostic shared DEGs of UCEC and PCOS were based on the Cox regression model. (A) The partial likelihood bias and log λ of lasso Coss analysis, and Lasso regression identified 16 candidate genes. (B) The coefficient and log λ of lasso Coss analysis. (C) The forest plot of nine prognostic shared DEGs was based on the multivariate Cox regression model.

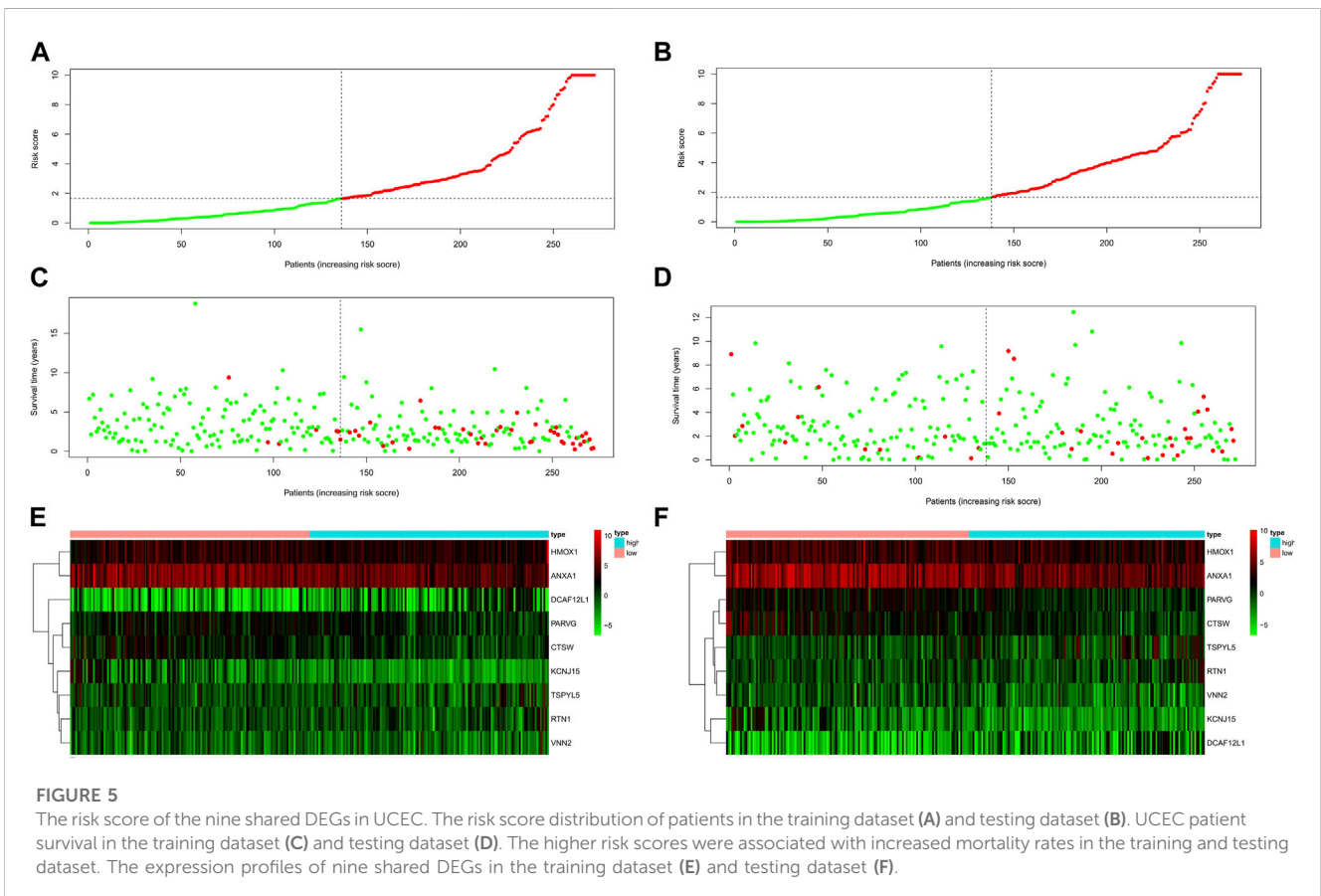
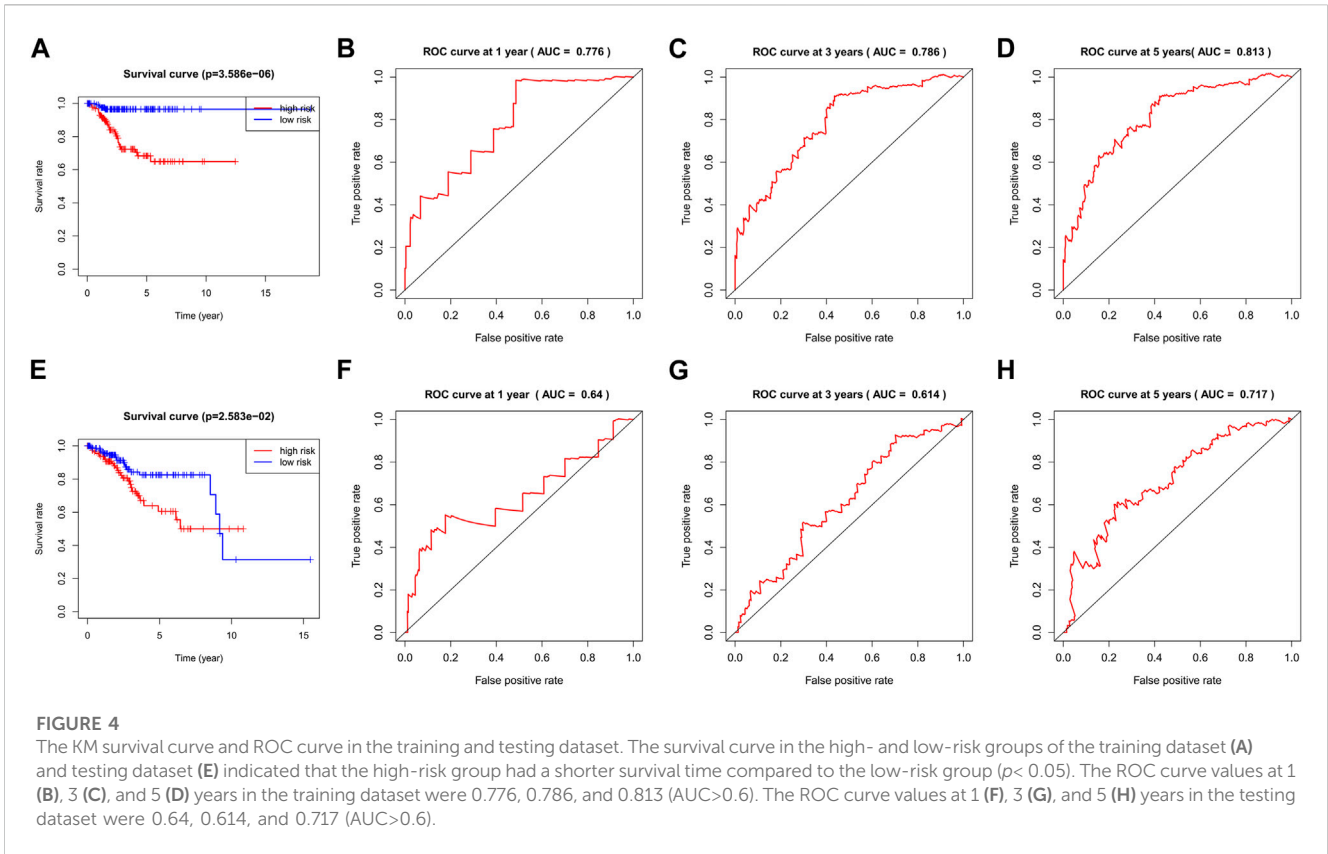
UCEC DEGs are presented in the heat map (Figure 1A). The analysis of PCOS dataset GSE34526 from the GEO database was performed using the Affymetrix Human Genome U133 Plus 2.0 Array platform, resulting in the identification of 2199 DEGs meeting the conditions, including 1373 upregulated differential genes and 826 downregulated differential genes. The PCOS DEGs are shown in the heat map (Figure 1B).

3.2 Confirming the shared DEGs for UCEC and PCOS

The Venn diagram demonstrates the presence of 151 shared DEGs between UCEC and PCOS (Figure 1C). Among the

TABLE 2 Multivariate Cox regression analyses of nine shared DEGs.

ID	Coef	HR	HR.95L	HR.95H	p-value
TSPYL5	0.033701	1.034275	1.008409	1.060806	0.009109
PARVG	-0.10052	0.904369	0.800222	1.022071	0.107344
KCNJ15	-0.67959	0.506825	0.27009	0.951057	0.034326
RTN1	0.200334	1.221811	1.115558	1.338183	1.59E-05
CTSW	-0.19341	0.824144	0.700535	0.969563	0.01966
HMOX1	0.002834	1.002838	1.001585	1.004093	8.97E-06
DCAF12L1	0.088495	1.092529	0.990085	1.205574	0.078136
VNN2	0.039031	1.039803	1.0096	1.07091	0.009454
ANXA1	-0.00733	0.992698	0.985906	0.999537	0.03643



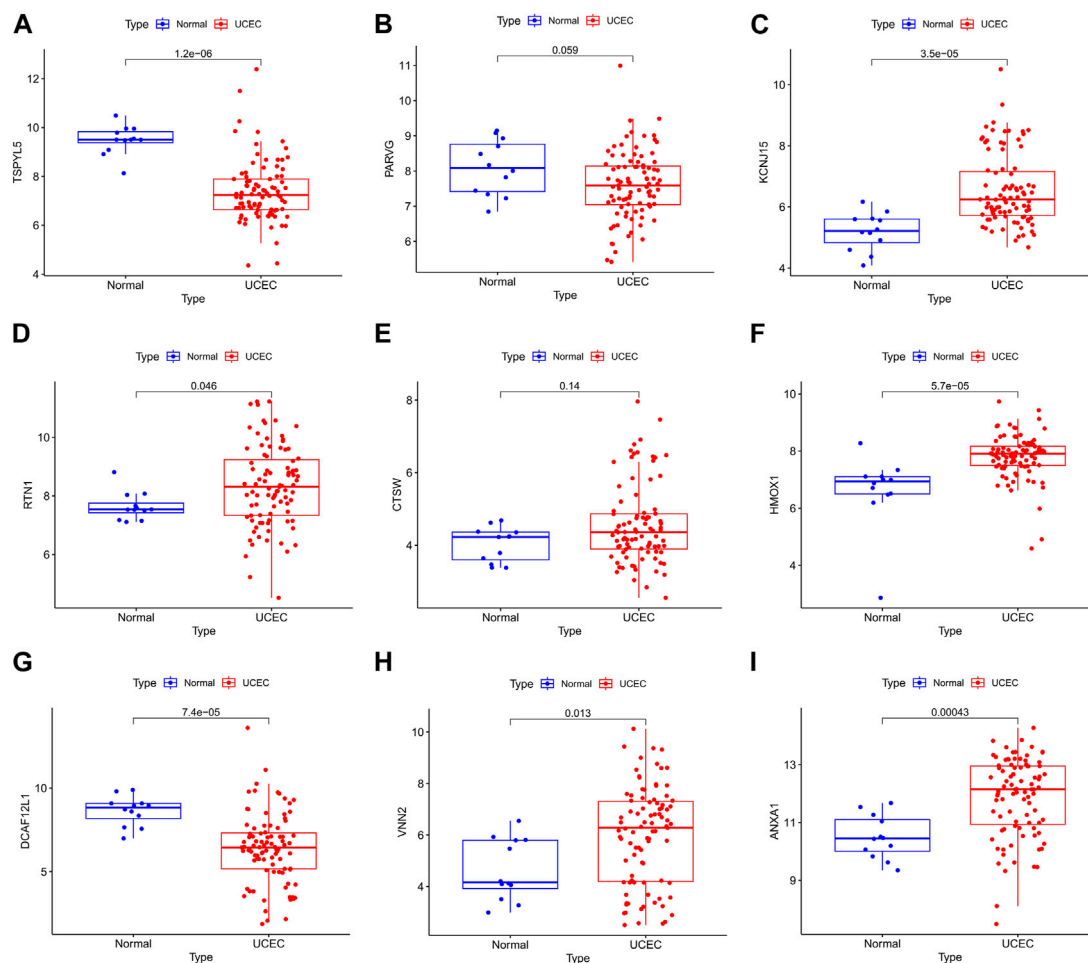


FIGURE 6

The bar chart of the nine DEGs in UCEC external verification. The expressions of TSPYL5 (A), KCN15 (C), RTN1 (D), HMOX1 (F), DCAF12L1 (G), VNN2 (H), and ANXA1 (I) were significantly different in normal control and UCEC ($p > 0.05$). The expressions of PARVG (B) and CTSW (E) were not different in normal control and UCEC ($p > 0.05$).

151 shared differential genes, 73 were upregulated and 78 were downregulated in UCEC (Supplementary Figure S1).

3.3 PPI network and enrichment analysis of shared DEGs for UCEC and PCOS

Based on the interactions of shared DEGs obtained from the Sting database, Cytoscape software was used to visualize the key shared DEGs (Figure 2A) and the PPI network diagram of shared genes (Figure 2B). It was determined that the top ten genes based on degree were LCP2, NKG7, IL2RG, CD7, CCR5, RAC2, CD79A, KIT, CCR7, and GBP5 (Table 1; Figures 2A, B). The shared DEGs were enriched using the Metascape online database, and PPI networks were based on cluster analysis (Figure 2C) and p -value (Figure 2D). The top three enrichment analyses were leukocyte activation, inflammatory response, and regulation of lymphocyte activation as the most significant processes (Figure 2E).

3.4 Establishment of shared DEGs prediction model for UCEC

The analysis of the 151 shared DEGs between UCEC and PCOS involved performing univariate Cox regression, which revealed 23 DEGs associated with UCEC survival. The lasso regression lambda optimal algorithm was employed, and tenfold cross-validation identified an optimal lambda value of -3.6 (Figure 3A). Utilizing lasso regression, 16 candidate genes associated with UCEC prognosis were identified, each having non-zero lasso coefficients (Figure 3B). Those genes include TSPYL5, ZNF683, SPOCK3, PARVG, DCAF12L2, KCN15, RTN1, NEURL1, SKAP2, KCNH2, ACKR3, CTSW, HMOX1, DCAF12L1, VNN2, and ANXA1. Subsequently, a multivariate Cox regression model was constructed, resulting in the identification of nine prognostic markers for UCEC. These markers are TSPYL5, PARVG, KCN15, RTN1, CTSW, HMOX1, DCAF12L1, VNN2, and ANXA1 (Table 2; Figure 3C). Among them, VNN2, PARVG, CTSW, HMOX1, and ANXA1 are upregulated in

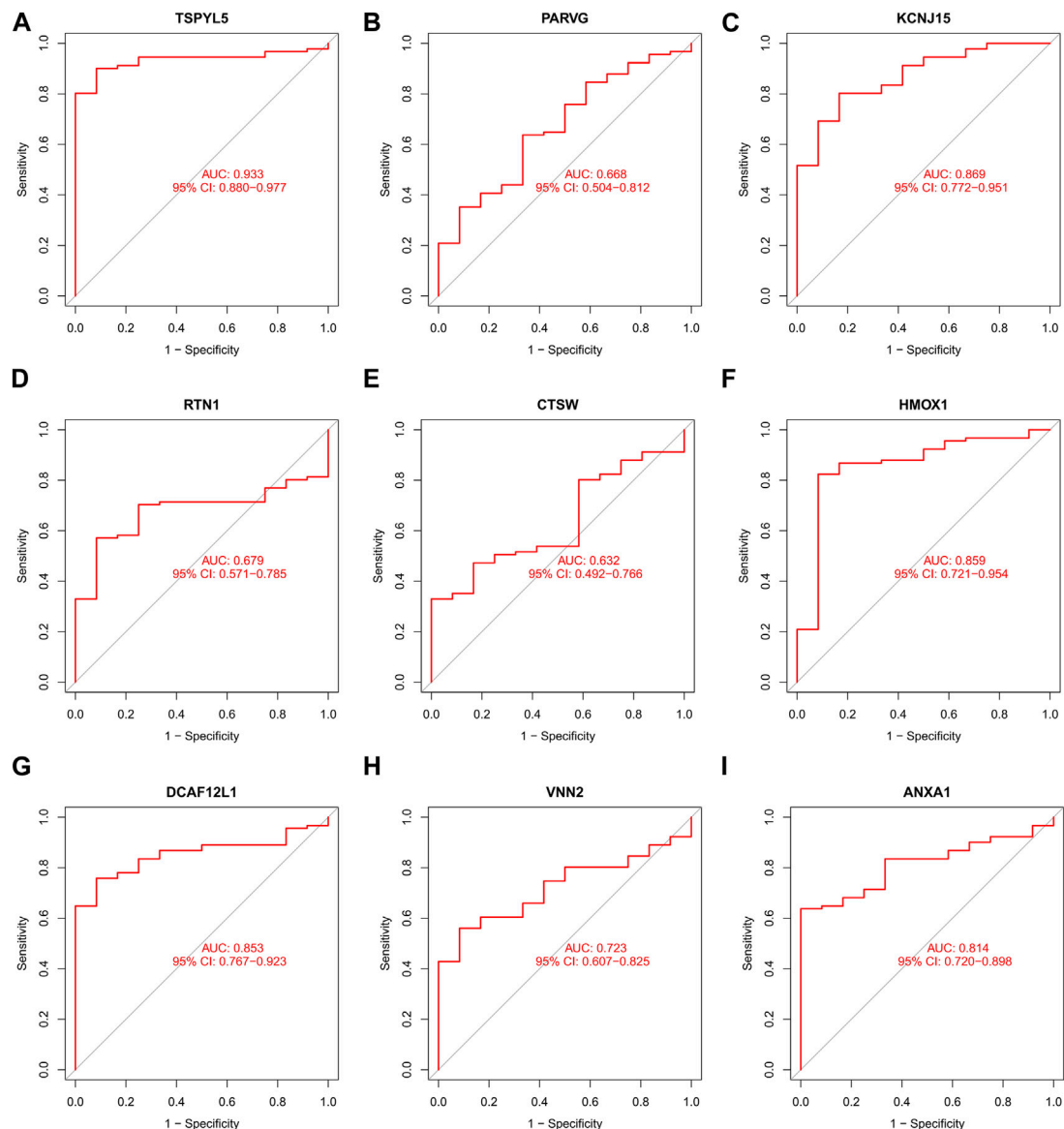


FIGURE 7

The ROC curve of the nine DEGs in UCEC external verification. The ROC curves of those nine shared DEGs had good predictive performances. The AUCs of TSPYL5 (A), PARVG (B), KCNJ15 (C), RTN1 (D), CTSW (E), HMOX1 (F), DCAF12L1 (G), VNN2 (H), and ANXA1 (I) were 0.933, 0.668, 0.869, 0.679, 0.632, 0.859, 0.853, 0.723, and 0.814, respectively.

UCEC, while DCAF12L1, KCNJ15, TSPYL5, and RTN1 are downregulated in UCEC (Supplementary Figure S2).

3.5 Internal validation of shared DEGs prediction models

The TCGA-UCEC data were divided into a training dataset (272 cases) and a testing dataset (272 cases). The median risk value of the training dataset was determined to be 1.653732. Using this median risk value as the cutoff, both the training datasets and testing datasets were divided into two groups: high-risk and low-risk. Each group consisted of 136 UCEC cases. The risk value proved to be a reliable predictor. The KM survival curves for the high-risk and low-risk groups in both the training and testing datasets indicated that

the high-risk group had a shorter survival time compared to the low-risk group ($p < 0.05$, Figures 4A, E).

The ROC curve values at 1, 3, and 5 years in the training dataset were 0.776, 0.786, and 0.813, respectively, (Figures 4B–D). The testing dataset was used to validate the model and the corresponding ROC curve values at 1, 3, and 5 years in the testing dataset were 0.64, 0.614, and 0.717, respectively, (Figures 4F–H). These ROC curve values indicate that the model constructed by the nine prognostic genes in both the UCEC training dataset and the testing dataset performed well, as all values were greater than 0.6.

The survival scores and status of the UCEC training dataset indicated that higher risk scores were associated with increased mortality rates (Figures 5A, C). Furthermore, when validating the risk score of the model using the testing dataset, the survival scores and status of the high-risk and low-risk groups also demonstrated that higher-risk scores were

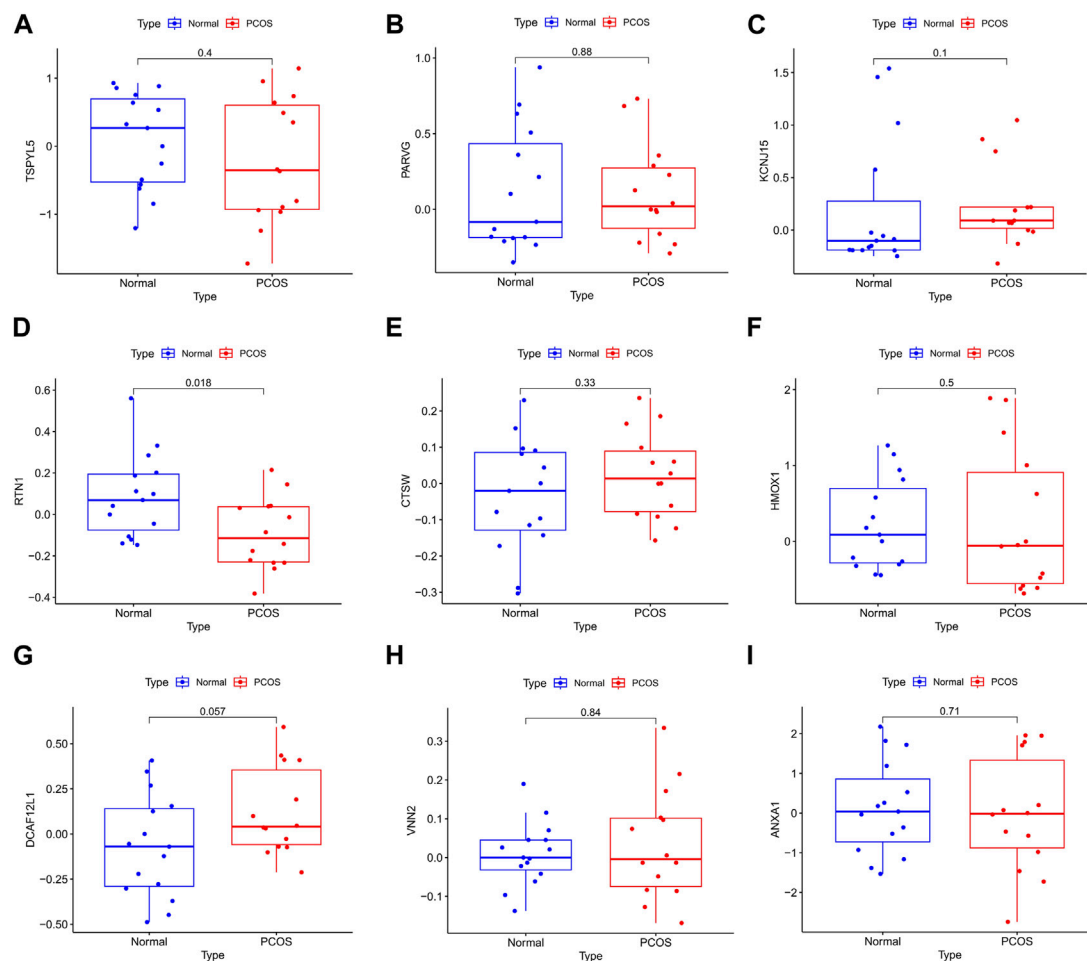


FIGURE 8

The bar chart of the nine DEGs in PCOS external verification. The expression of RTN1 (D) was significantly different in normal control and PCOS ($p < 0.05$). The expressions of TSPYL5 (A), PARVG (B), KCNJ15 (C), CTSW (E), HMOX1 (F), DCAF12L1 (G), VNN2 (H), and ANXA1 (I) were not different in normal control and PCOS ($p > 0.05$).

correlated with higher mortality rates (Figures 5B, D). The expression levels of the nine key shared DEGs in both the training and testing datasets were visualized using a heat map (Figures 5E, F).

3.6 External validation of prognostic genes

The UCEC dataset GSE17025 was used as an external validation dataset to validate the prognostic genes TSPYL5, PARVG, KCNJ15, RTN1, CTSW, HMOX1, DCAF12L1, VNN2, and ANXA1. Among them, the expressions of TSPYL5, KCNJ15, RTN1, HMOX1, DCAF12L1, VNN2, and ANXA1 exhibited significant differences between normal controls and UCEC ($p < 0.05$, Figure 6), confirming their status as key prognostic genes for UCEC. The ROC curves for all nine prognostic genes in the UCEC external validation set were over 0.6 (Figure 7A–I), indicating their good predictive performance. The PCOS dataset GSE48301 was used as an external validation dataset to validate the same nine prognostic genes. The results showed that RTN1 exhibited a

significant difference between normal controls and PCOS ($p < 0.05$, Figure 8D). However, the expressions of TSPYL5 (Figure 8A), PARVG (Figure 8B), KCNJ15 (Figure 8C), CTSW (Figure 8E), HMOX1 (Figure 8F), DCAF12L1 (Figure 8G), VNN2 (Figure 8H), and ANXA1 (Figure 8I) were not different in normal control and PCOS ($p > 0.05$). Moreover, the ROC curves for KCNJ15 (Figure 9C), RTN1 (Figure 9D), CTSW (Figure 9E), and DCAF12L1 (Figure 9G) demonstrated good predictive performance ($AUC > 0.6$). Therefore, RTN1 was identified as a potential marker in both UCEC and PCOS.

3.7 RFS and expression analysis by Gepia

The expressions of the key prognostic genes TSPYL5, KCNJ15, RTN1, HMOX1, DCAF12L1, VNN2, and ANXA1 in UCEC were analyzed using Gepia. The results showed that higher expression of RTN1 was associated with longer RFS in UCEC patients ($p < 0.05$, Figure 10C). However, the expressions of TSPYL5 (Figure 10A), KCNJ15 (Figure 10B), HMOX1 (Figure 10D), DCAF12L1

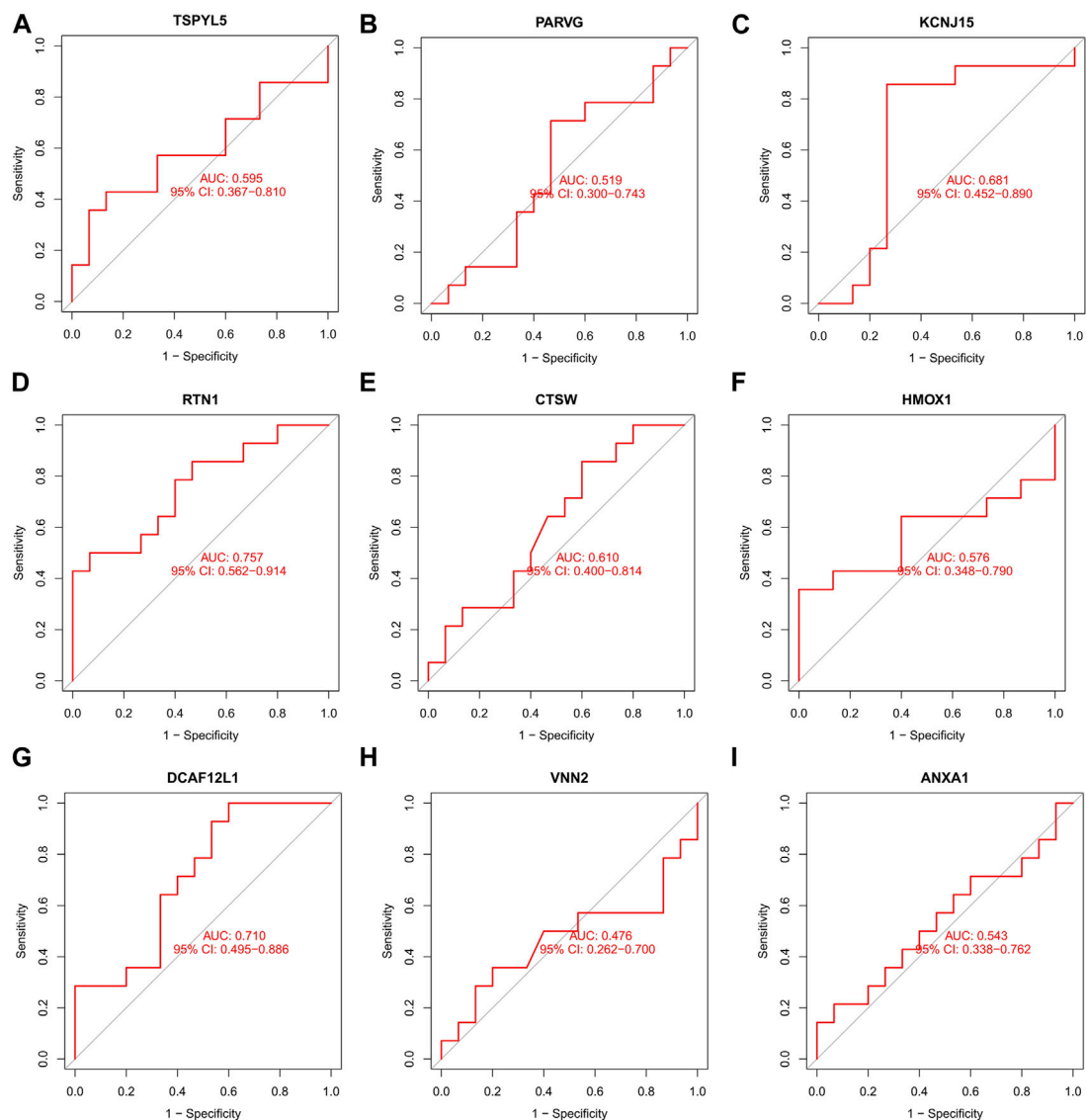


FIGURE 9

The ROC curve of the nine DEGs in PCOS external verification. The ROC curves of KCN15 (C), RTN1 (D), CTSW (E), and DCAF12L1 (G) were 0.681, 0.757, 0.610, and 0.710, respectively, indicating good predictive performance (AUC>0.6). The ROC curves of TSPYL5 (A), PARVG (B), KCN15 (C), HMOX1 (F), VNN2 (H), and ANXA1 (I) had no predictive performances (AUC<0.6).

(Figure 10E), VNN2 (Figure 10F), and ANXA1 (Figure 10G) were not found to be related to the RFS of UCEC patients ($p > 0.05$).

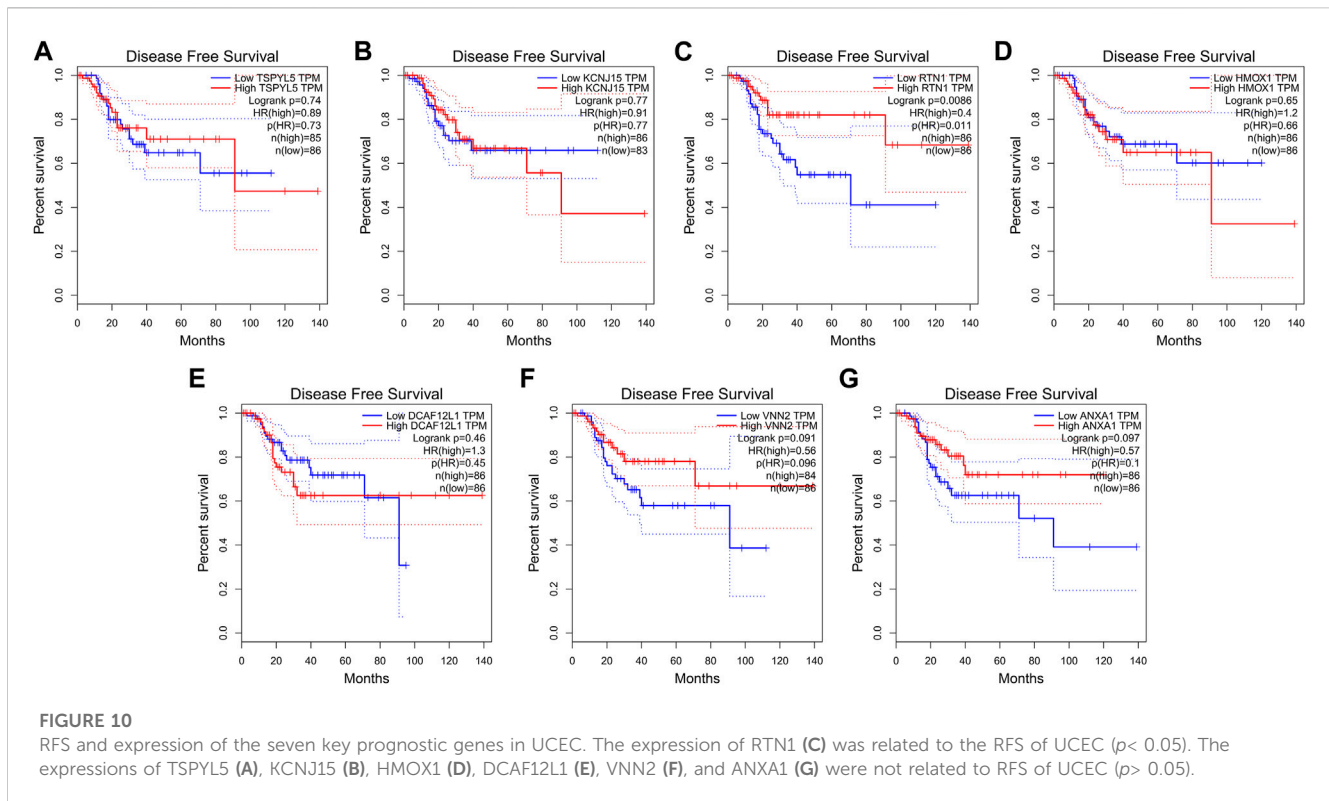
3.8 Immune infiltration analysis of shared DEGs

Based on the Cibersort algorithm, we analyzed the infiltration of immune cells in UCEC and PCOS using the shared DEGs. Figure 11A shows the proportion of 22 immune cell subsets in the 151 shared DEGs. We processed the shared DEGs using R software and stratified them into high and low TMB groups. The results indicated that T cells CD8, T cells CD4 memory resting, T cells CD4 memory activated, T cells follicular helper, T cells regulatory, NK cells activated, macrophages M1, macrophages M2,

mast cells resting, and mast cells activated exhibited significantly different expressions between the UCEC high and low TMB groups ($p < 0.05$, Figure 11B).

3.9 Regulatory relationship between key prognostic genes and immune cells in UCEC

The key prognostic genes TSPYL5, KCN15, RTN1, HMOX1, DCAF12L1, VNN2, and ANXA1 in UCEC were found to have regulatory relationships with immune cells such as T cells, B cells, NK cells, and others (Figures 12A–G). Among them, RTN1 was positively correlated with resting dendritic cells, resting CD4 memory T cells, neutrophils, and M0 macrophages, and negatively correlated with follicular helper T cells. (Figure 12C).



4 Discussion

UCEC, the fourth most prevalent female malignancy worldwide, is a common tumor in the female reproductive system (Cai et al., 2021). Its incidence is increasing globally, especially in developed regions (Makker et al., 2021; Banz-Jansen et al., 2022; Crosbie et al., 2022), which can be attributed to the rising prevalence of obesity (Rahib et al., 2014; Lauby-Secretan et al., 2016). UCEC has a low survival rate and poor prognosis (Brooks et al., 2019; Liu et al., 2021a), with a 5-year overall survival rate of 15%–17% for recurrent UCEC (Makker et al., 2021). However, early diagnosis and treatment of UCEC lead to better outcomes (Oaknin et al., 2022). Therefore, early diagnosis and treatment are crucial for UCEC. PCOS is a common disorder among women of reproductive age, with a global prevalence ranging from 6% to 20% (Joham et al., 2021; Kumariya et al., 2021). Obesity and low-grade inflammation associated with obesity are common features of PCOS (Barrea et al., 2021). Women with PCOS have an increased risk of developing tumors, including UCEC, ovarian cancer, breast cancer, and others (Gadducci et al., 2005; Xu et al., 2017; Meczekalski et al., 2020; Li Z. et al., 2022; Mitric and Bernardini, 2022). However, clinical outcomes of patients with the same tumor stage vary considerably, suggesting that conventional staging techniques may not accurately predict cancer prognosis (Xu et al., 2017). This emphasizes the importance of genetic diagnosis and therapy. Therefore, our focus is to explore the pathophysiological association between UCEC and PCOS and identify potential genetic biomarkers of UCEC.

Our study identified 1949 DEGs in UCEC, 2199 DEGs in PCOS, and 151 shared DEGs in UCEC and PCOS when compared

to normal tissues. The shared DEGs PPI network of UCEC and PCOS was visualized using Cytoscape software, and Metascape-generated PPI networks were used for gene enrichment analysis (Li et al., 2019). GO and KEGG enrichment analysis of the 151 shared DEGs through the Metascape online database revealed that they were primarily enriched in leukocyte activation. Through model construction and internal verification using the TCGA database, TSPYL5, PARVG, KCNJ15, RTN1, CTSW, HMOX1, DCAF12L1, VNN2, and ANXA1 were identified as prognostic genes for UCEC within the shared DEGs of UCEC and PCOS. Additionally, the UCEC high-risk group exhibited a shorter survival time and worse prognosis compared to the low-risk group. External validation with additional datasets confirmed that TSPYL5, KCNJ15, RTN1, HMOX1, DCAF12L1, VNN2, and ANXA1 were key prognostic genes for UCEC. Analysis of these seven key prognostic genes using Gepia revealed that higher expression of RTN1 was associated with longer RFS in UCEC patients, indicating that detecting RTN1 expression may aid in the prognosis assessment of UCEC patients. Furthermore, RTN1 was identified as a potential marker in both UCEC and PCOS. Several studies have indicated that the seven key prognostic genes identified in this study for UCEC, including TSPYL5, VNN2, ANXA1, and DCAF12L1, are associated with tumor occurrence and development (Feng et al., 2020; Soler et al., 2020; Hernandez-Meza et al., 2021; Lu et al., 2021). In particular, DCAF12L1 has been considered a methylated gene contributing to the development of UCEC (Lu et al., 2021). Moreover, diabetes is a risk factor in both PCOS and UCEC (Lauby-Secretan et al., 2016; Patel, 2018; Liao et al., 2021), and HMOX1, RTN1, and KCNJ15 are considered diabetes-related genes (Okamoto et al., 2012; Meng et al., 2021; Meng et al., 2022).

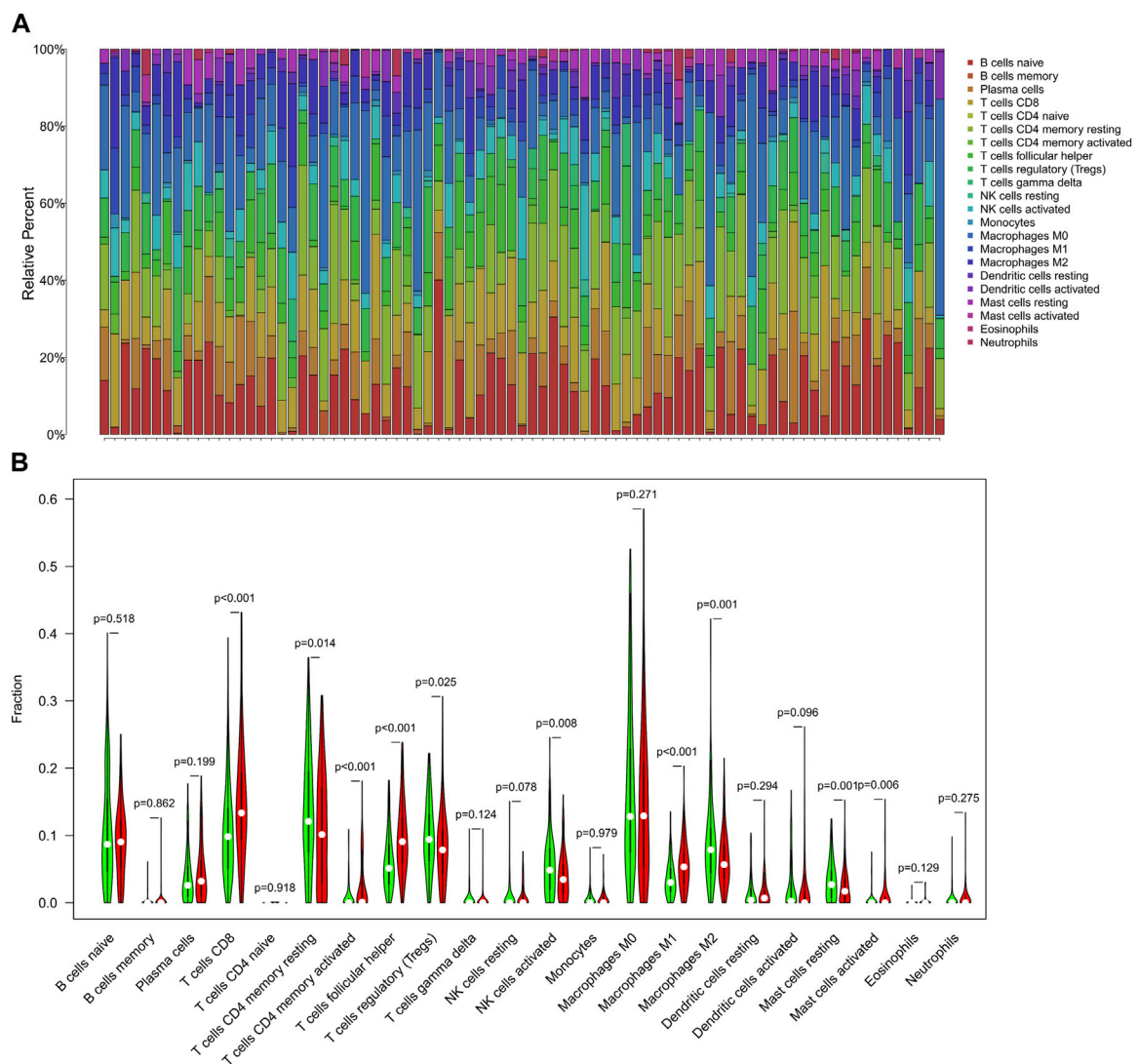
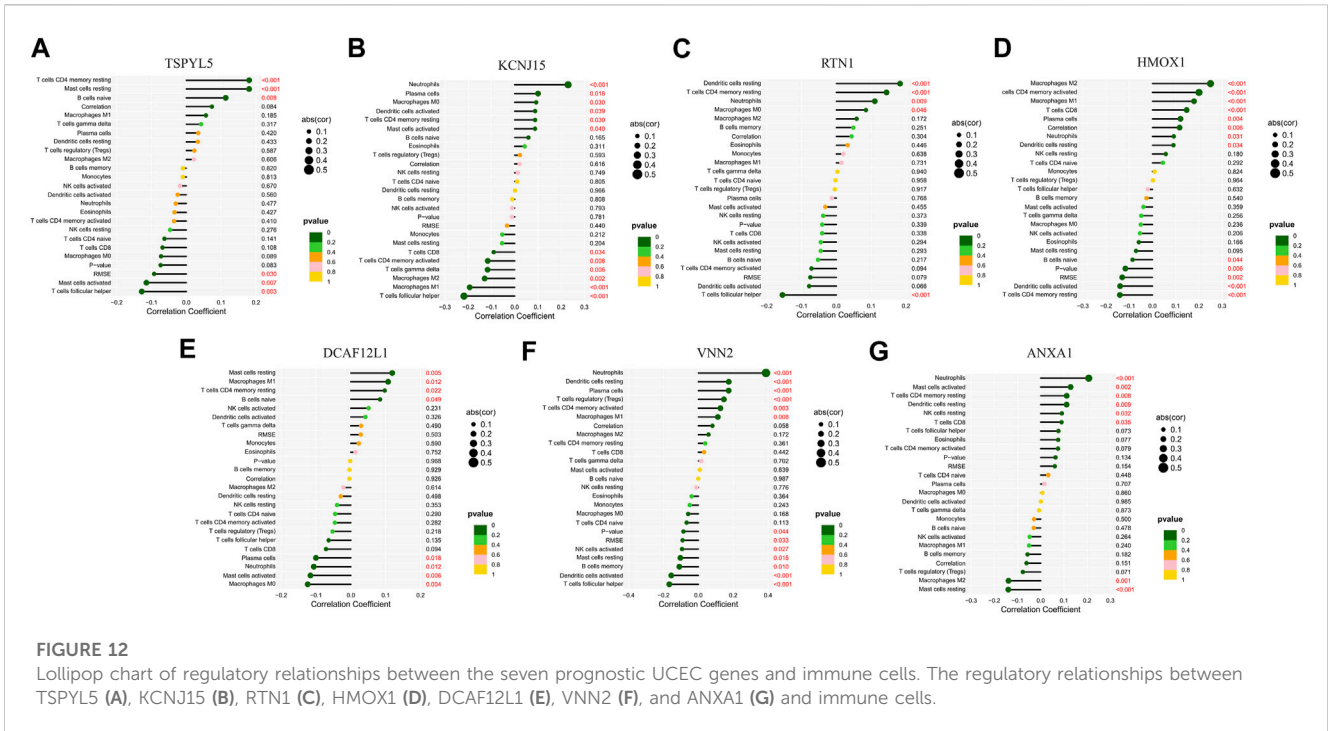


FIGURE 11

The immune infiltration analysis of shared DEGs. **(A)** The bar chart of shared DEGs and immune cell subsets. Each column represents a sample, and different colors represent different immune cells. **(B)** The violin diagram of the TMB in shared DEGs. The Wilcoxon rank-sum test showed that T cells CD8 ($p < 0.001$), T cells CD4 memory resting ($p = 0.014$), T cells CD4 memory activated ($p < 0.001$), T cells follicular helper ($p < 0.001$), T cells regulatory ($p = 0.025$), NK cells activated ($p = 0.008$), Macrophages M1 ($p < 0.001$), Macrophages M2 ($p = 0.001$), Mast cells resting ($p = 0.001$), and Mast cells activated ($p = 0.006$) had a difference in high and low TMB groups.

Great progress has been made in the treatment of UCEC, but advanced or recurrent patients still present treatment challenges (Zhang Y. et al., 2022). Immunotherapy is a promising treatment approach that utilizes drugs to enhance the immune cells' infiltration in the tumor microenvironment and activate the immune system to recognize and attack cancer cells (O'Donnell et al., 2019; Tao et al., 2021; Li X. et al., 2022; Marin-Jimenez et al., 2022; Zhang Y. et al., 2022). This method can effectively reduce the damage to healthy cells caused by treatment (Marin-Jimenez et al., 2022). Checkpoint inhibitor immunotherapy has revolutionized the treatment of various tumors, including UCEC (Marin-Jimenez et al., 2022). Pembrolizumab, a checkpoint inhibitor, has demonstrated efficacy in treating tumors with TMB, and pembrolizumab/lenvatinib has been used in the treatment of UCEC with high

TMB (Marin-Jimenez et al., 2022). The results of this study on the immune infiltration of UCEC and PCOS-shared DEGs revealed differences in the expression of immune cells such as NK cells, CD8⁺ T cells, macrophages, and mast cells between patients with high and low TMB in UCEC. Further analysis of the seven key prognostic genes of UCEC indicated regulatory relationships with immune cells. However, the main limitation of this study is the lack of clinical validation experiments, as only samples from the TCGA database, GSE34526 dataset, GSE17025 dataset, and GSE48301 dataset were included. Although the altered expressions of TSPYL5, KCN115, RTN1, HMOX1, DCAF12L1, VNN2, and ANXA1 are associated with UCEC survival, TMB, immune infiltration, and immune cell regulation, it does not necessarily imply that the risk prediction model can be applied in actual clinical practice to provide a basis for immunotherapy.



In summary, this study successfully developed and validated a potential marker model for UCEC and PCOS, which identified TSPYL5, PARVG, KCNJ15, RTN1, CTSW, HMOX1, DCAF12L1, VNN2, and ANXA1 as prognostic genes for UCEC. Among these genes, TSPYL5, KCNJ15, RTN1, HMOX1, DCAF12L1, VNN2, and ANXA1 were found to be associated with the survival time, TMB, and immune infiltration in UCEC. RTN1 may serve as a novel immunotherapy biomarker for both UCEC and PCOS.

Data availability statement

The datasets GSE34526, GSE17025, and GSE48301 for this study can be found in the GEO database <https://www.ncbi.nlm.nih.gov/geo/>. The dataset UCEC for this study can be found in the TCGA database <https://portal.gdc.cancer.gov/>.

Author contributions

RW and WZ designed and conceived the study. RW, CW, and BZ collect and analyze data under the guidance of WZ and JL. WZ and JL oversaw the research. RW and CW wrote the first draft of the article, and JL revised the article. All authors contributed to the article and approved the submitted version.

References

Atsavaprane, E. S., Billingsley, M. M., and Mitchell, M. J. (2021). Delivery technologies for T cell gene editing: Applications in cancer immunotherapy. *EBioMedicine* 67, 103354. doi:10.1016/j.ebiom.2021.103354

Acknowledgments

We would like to gratefully acknowledge contributions from TCGA and GEO.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmolb.2023.1192313/full#supplementary-material>

Banz-Jansen, C., Helweg, L. P., and Kaltschmidt, B. (2022). Endometrial cancer stem cells: Where do we stand and where should we go? *Int. J. Mol. Sci.* 23 (6), 3412. doi:10.3390/ijms23063412

- Barrea, L., Muscogiuri, G., Pugliese, G., de Alteriis, G., Colao, A., and Savastano, S. (2021). Metabolically Healthy obesity (MHO) vs. Metabolically unhealthy obesity (MUO) phenotypes in PCOS: Association with endocrine-metabolic profile, adherence to the Mediterranean diet, and body composition. *Nutrients* 13 (11), 3925. doi:10.3390/nu13113925
- Barry, J. A., Azizia, M. M., and Hardiman, P. J. (2014). Risk of endometrial, ovarian and breast cancer in women with polycystic ovary syndrome: A systematic review and meta-analysis. *Hum. Reprod. Update* 20 (5), 748–758. doi:10.1093/humupd/dmu012
- Brooks, R. A., Fleming, G. F., Lastra, R. R., Lee, N. K., Moroney, J. W., Son, C. H., et al. (2019). Current recommendations and recent progress in endometrial cancer. *CA Cancer J. Clin.* 69 (4), 258–279. doi:10.3322/caac.21561
- Cai, Y., Wang, B., Xu, W., Liu, K., Gao, Y., Guo, C., et al. (2021). Endometrial cancer: Genetic, metabolic characteristics, therapeutic strategies and nanomedicine. *Curr. Med. Chem.* 28 (42), 8755–8781. doi:10.2174/09298673288666210705144456
- Crosbie, E. J., Kitson, S. J., McAlpine, J. N., Mukhopadhyay, A., Powell, M. E., and Singh, N. (2022). Endometrial cancer. *Lancet* 399 (10333), 1412–1428. doi:10.1016/S0140-6736(22)00323-3
- El-Mayta, R., Zhang, Z., Hamilton, A. G., and Mitchell, M. J. (2021). Delivery technologies to engineer natural killer cells for cancer immunotherapy. *Cancer Gene Ther.* 28 (9), 947–959. doi:10.1038/s41417-021-00336-2
- Feng, J., Lu, S. S., Xiao, T., Huang, W., Yi, H., Zhu, W., et al. (2020). ANXA1 binds and stabilizes EphA2 to promote nasopharyngeal carcinoma growth and metastasis. *Cancer Res.* 80 (20), 4386–4398. doi:10.1158/0008-5472.CAN-20-0560
- Gadducci, A., Gargini, A., Palla, E., Fanucchi, A., and Genazzani, A. R. (2005). Polycystic ovary syndrome and gynecological cancers: Is there a link? *Gynecol. Endocrinol.* 20 (4), 200–208. doi:10.1080/09513590400021201
- Hernandez-Meza, G., von Felden, J., Gonzalez-Kozlova, E. E., Garcia-Lezana, T., Peix, J., Portela, A., et al. (2021). DNA methylation profiling of human hepatocarcinogenesis. *Hepatology* 74 (1), 183–199. doi:10.1002/hep.31659
- Joham, A. E., Kakoly, N. S., Teede, H. J., and Earnest, A. (2021). Incidence and predictors of hypertension in a cohort of Australian women with and without polycystic ovary syndrome. *J. Clin. Endocrinol. Metab.* 106 (6), 1585–1593. doi:10.1210/clinem/dgab134
- Kennedy, L. B., and Salama, A. K. S. (2020). A review of cancer immunotherapy toxicity. *CA Cancer J. Clin.* 70 (2), 86–104. doi:10.3322/caac.21596
- Kumariya, S., Ubba, V., Jha, R. K., and Gayen, J. R. (2021). Autophagy in ovary and polycystic ovary syndrome: Role, dispute and future perspective. *Autophagy* 17 (10), 2706–2733. doi:10.1080/15548627.2021.1938914
- Lauby-Secretan, B., Scoccianti, C., Loomis, D., Grosse, Y., Bianchini, F., Straif, K., et al. (2016). Body fatness and cancer—viewpoint of the IARC working group. *N. Engl. J. Med.* 375 (8), 794–798. doi:10.1056/NEJMs1606602
- Li, W., Wang, S., Qiu, C., Liu, Z., Zhou, Q., Kong, D., et al. (2019). Comprehensive bioinformatics analysis of acquired progesterone resistance in endometrial cancer cell line. *J. Transl. Med.* 17 (1), 58. doi:10.1186/s12967-019-1814-6
- Li, X., Zhou, J., Zhang, W., You, W., Wang, J., Zhou, L., et al. (2022a). Pan-cancer analysis identifies tumor cell surface targets for CAR-T cell therapies and antibody drug conjugates. *Cancers (Basel)* 14 (22), 5674. doi:10.3390/cancers14225674
- Li, Z., Wang, Y. H., Wang, L. L., Hu, D. T., Teng, Y., Zhang, T. Y., et al. (2022b). Polycystic ovary syndrome and the risk of endometrial, ovarian and breast cancer: An updated meta-analysis. *Scott Med. J.* 67 (3), 109–120. doi:10.1177/00369330221107099
- Liao, B., Qiao, J., and Pang, Y. (2021). Central regulation of PCOS: Abnormal neuronal-reproductive-metabolic circuits in PCOS pathophysiology. *Front. Endocrinol. (Lausanne)* 12, 667422. doi:10.3389/fendo.2021.667422
- Liu, J., Geng, R., Yang, S., Shao, F., Zhong, Z., Yang, M., et al. (2021a). Development and clinical validation of novel 8-gene prognostic signature associated with the proportion of regulatory T cells by weighted gene Co-expression network analysis in uterine corpus endometrial carcinoma. *Front. Immunol.* 12, 788431. doi:10.3389/fimmu.2021.788431
- Liu, J., Ji, C., Wang, Y., Zhang, C., and Zhu, H. (2021b). Identification of methylation-driven genes prognosis signature and immune microenvironment in uterine corpus endometrial cancer. *Cancer Cell Int.* 21 (1), 365. doi:10.1186/s12935-021-02038-z
- Lu, Y., Tang, W., Wang, X., Kang, X., You, J., and Chen, L. (2021). Development of potential prognostic biomarkers based on DNA methylation-driven genes for patients with endometrial cancer. *Int. J. Gen. Med.* 14, 10541–10555. doi:10.2147/IJGM.S341771
- Makker, V., MacKay, H., Ray-Coquard, L., Levine, D. A., Westin, S. N., Aoki, D., et al. (2021). Endometrial cancer. *Nat. Rev. Dis. Prim.* 7 (1), 88. doi:10.1038/s41572-021-00324-8
- Makker, V., Rasco, D., Vogelzang, N. J., Brose, M. S., Cohn, A. L., Mier, J., et al. (2019). Lenvatinib plus pembrolizumab in patients with advanced endometrial cancer: An interim analysis of a multicentre, open-label, single-arm, phase 2 trial. *Lancet Oncol.* 20 (5), 711–718. doi:10.1016/S1470-2045(19)30020-8
- Marin-Jimenez, J. A., Garcia-Mulero, S., Matias-Guiu, X., and Piuals, J. M. (2022). Facts and hopes in immunotherapy of endometrial cancer. *Clin. Cancer Res.* 28 (22), 4849–4860. doi:10.1158/1078-0432.CCR-21-1564
- Meczekalski, B., Perez-Roncero, G. R., Lopez-Baena, M. T., Chedraui, P., and Perez-Lopez, F. R. (2020). The polycystic ovary syndrome and gynecological cancer risk. *Gynecol. Endocrinol.* 36 (4), 289–293. doi:10.1080/09513590.2020.1730794
- Meng, D., Wu, L., Li, Z., Ma, X., Zhao, S., Zhao, D., et al. (2022). LncRNA TUG1 ameliorates diabetic nephropathy via inhibition of PU.1/RTN1 signaling pathway. *J. Leukoc. Biol.* 111 (3), 553–562. doi:10.1002/JLB.6A1020-699RRR
- Meng, Z., Liang, H., Zhao, J., Gao, J., Liu, C., Ma, X., et al. (2021). HMOX1 upregulation promotes ferroptosis in diabetic atherosclerosis. *Life Sci.* 284, 119935. doi:10.1016/j.lfs.2021.119935
- Mitric, C., and Bernardini, M. Q. (2022). Endometrial cancer: Transitioning from histology to genomics. *Curr. Oncol.* 29 (2), 741–757. doi:10.3390/curroncol29020063
- O'Donnell, J. S., Teng, M. W. L., and Smyth, M. J. (2019). Cancer immunoeediting and resistance to T cell-based immunotherapy. *Nat. Rev. Clin. Oncol.* 16 (3), 151–167. doi:10.1038/s41571-018-0142-8
- Oaknin, A., Bosse, T. J., Creutzberg, C. L., Giornelli, G., Harter, P., Joly, F., et al. (2022). Endometrial cancer: ESMO clinical practice guideline for diagnosis, treatment and follow-up. *Ann. Oncol.* 33 (9), 860–877. doi:10.1016/j.annonc.2022.05.009
- Okamoto, K., Iwasaki, N., Doi, K., Noiri, E., Iwamoto, Y., Uchigata, Y., et al. (2012). Inhibition of glucose-stimulated insulin secretion by KCNJ15, a newly identified susceptibility gene for type 2 diabetes. *Diabetes* 61 (7), 1734–1741. doi:10.2337/db11-1201
- Patel, S. (2018). Polycystic ovary syndrome (PCOS), an inflammatory, systemic, lifestyle endocrinopathy. *J. Steroid Biochem. Mol. Biol.* 182, 27–36. doi:10.1016/j.jsbmb.2018.04.008
- Rahib, L., Smith, B. D., Aizenberg, R., Rosenzweig, A. B., Fleshman, J. M., and Matrisian, L. M. (2014). Projecting cancer incidence and deaths to 2030: The unexpected burden of thyroid, liver, and pancreas cancers in the United States. *Cancer Res.* 74 (11), 2913–2921. doi:10.1158/0008-5472.CAN-14-0155
- Rego, S. M., and Snyder, M. P. (2019). High throughput sequencing and assessing disease risk. *Cold Spring Harb. Perspect. Med.* 9 (1), a026849. doi:10.1101/cshperspect.a026849
- Riley, R. S., June, C. H., Langer, R., and Mitchell, M. J. (2019). Delivery technologies for cancer immunotherapy. *Nat. Rev. Drug Discov.* 18 (3), 175–196. doi:10.1038/s41573-018-0006-z
- Rodriguez, R., and Miller, K. M. (2014). Unravelling the genomic targets of small molecules using high-throughput sequencing. *Nat. Rev. Genet.* 15 (12), 783–796. doi:10.1038/nrg3796
- Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13 (11), 2498–2504. doi:10.1101/gr.1239303
- Soler, D. C., Kerstetter-Fogle, A., Elder, T., Raghavan, A., Barnholtz-Sloan, J. S., Cooper, K. D., et al. (2020). A liquid biopsy to assess brain tumor recurrence: Presence of circulating Mo-mdsc and CD14+ VNN2+ myeloid cells as biomarkers that distinguish brain metastasis from radiation necrosis following stereotactic radiosurgery. *Neurosurgery* 88 (1), E67–E72. doi:10.1093/neuros/nyaa334
- Szklarczyk, D., Kirsch, R., Koutrouli, M., Nastou, K., Mehryary, F., Hachilif, R., et al. (2023). The STRING database in 2023: Protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Res.* 51 (D1), D638–D646. doi:10.1093/nar/gkac1000
- Szklarczyk, D., Morris, J. H., Cook, H., Kuhn, M., Wyder, S., Simonovic, M., et al. (2017). The STRING database in 2017: Quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Res.* 45 (D1), D362–D368. doi:10.1093/nar/gkw937
- Tang, Y., Li, M., Wang, J., Pan, Y., and Wu, F. X. (2015). CytoNCA: A cytoscape plugin for centrality analysis and evaluation of protein interaction networks. *Biosystems* 127, 67–72. doi:10.1016/j.biosystems.2014.11.005
- Tang, Z., Li, C., Kang, B., Gao, G., Li, C., and Zhang, Z. (2017). Gepia: A web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic Acids Res.* 45 (W1), W98–W102. doi:10.1093/nar/gkx247
- Tao, Y., Shen, H., Liu, Y., Li, G., Huang, Z., and Liu, Y. (2021). IL-23R in laryngeal cancer: A cancer immunoeediting process that facilitates tumor cell proliferation and results in cisplatin resistance. *Carcinogenesis* 42 (1), 118–126. doi:10.1093/carcin/bgaa058
- von Mering, C., Huynen, M., Jaeggi, D., Schmidt, S., Bork, P., and Snel, B. (2003). STRING: A database of predicted functional associations between proteins. *Nucleic Acids Res.* 31 (1), 258–261. doi:10.1093/nar/gkg034
- Xu, G., Zhang, M., Zhu, H., and Xu, J. (2017). A 15-gene signature for prediction of colon cancer recurrence and prognosis based on SVM. *Gene* 604, 33–40. doi:10.1016/j.gene.2016.12.016
- Zhang, M., Liu, Y., Hou, S., Wang, Y., Wang, C., Yin, Y., et al. (2022a). KDM4B, a potential prognostic biomarker revealed by large-scale public databases and clinical samples in uterine corpus endometrial carcinoma. *Mol. Omics* 18 (6), 506–519. doi:10.1039/d1mo00287b
- Zhang, Y., Yang, R., Xu, C., Zhang, Y., Deng, M., Wu, D., et al. (2022b). Analysis of the immune checkpoint lymphocyte activation gene-3 (LAG-3) in endometrial cancer: An emerging target for immunotherapy. *Pathol. Res. Pract.* 236, 153990. doi:10.1016/j.prp.2022.153990
- Zhou, Y., Zhou, B., Pache, L., Chang, M., Khodabakhshi, A. H., Tanaseichuk, O., et al. (2019). Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat. Commun.* 10 (1), 1523. doi:10.1038/s41467-019-09234-6