



Article

Diagnosis and Prediction of Endometrial Carcinoma Using Machine Learning and Artificial Neural Networks Based on Public Databases

Dongli Zhao ^{1,†}, Zhe Zhang ^{2,†}, Zhonghuang Wang ^{3,4}, Zhenglin Du ³ , Meng Wu ⁵, Tingting Zhang ¹, Jialu Zhou ¹, Wenming Zhao ^{3,4,*‡} and Yuanguang Meng ^{1,2,6,*‡}

¹ Department of Obstetrics & Gynecology, Chinese People's Liberation Army (PLA) Medical School, No. 28, Fuxing Road, Haidian District, Beijing 100853, China; zz18813066251@163.com (D.Z.); 17865190928@163.com (T.Z.); kalozzhou@163.com (J.Z.)

² Department of Obstetrics and Gynecology, Seventh Medical Center of Chinese PLA General Hospital, No. 5, Nanmencang, Dongshishitiao, Dongcheng District, Beijing 100700, China; zhangzhe301@126.com

³ National Genomics Data Center & CAS Key Laboratory of Genome Sciences and Information, Beijing Institute of Genomics, Chinese Academy of Sciences and China National Center for Bioinformation, Building 104, Courtyard 1, Beichen West Road, Chaoyang District, Beijing 100101, China; wangzhonghuang17m@big.ac.cn (Z.W.); duzhl@big.ac.cn (Z.D.)

⁴ University of Chinese Academy of Sciences, 19 Yuquan Road (a), Shijingshan District, Beijing 100049, China

⁵ Medical College, Graduate School of Nankai University, No. 94, Weijin Road, Nankai District, Tianjin 300110, China; awumeng_1992@163.com

⁶ Department of Gynecology and Obstetrics, Chinese PLA General Hospital, No. 28, Fuxing Road, Haidian District, Beijing 100853, China

* Correspondence: zhaowm@big.ac.cn (W.Z.); meng6512@vip.sina.com (Y.M.)

† These authors contributed equally to this work.

‡ These authors jointly supervised this work.



Citation: Zhao, D.; Zhang, Z.; Wang, Z.; Du, Z.; Wu, M.; Zhang, T.; Zhou, J.; Zhao, W.; Meng, Y. Diagnosis and Prediction of Endometrial Carcinoma Using Machine Learning and Artificial Neural Networks Based on Public Databases. *Genes* **2022**, *13*, 935. <https://doi.org/10.3390/genes13060935>

Academic Editors: Piero Fariselli and Peixin Dong

Received: 7 April 2022

Accepted: 20 May 2022

Published: 24 May 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: Endometrial carcinoma (EC), a common female reproductive system malignant tumor, affects thousands of people with high morbidity and mortality worldwide. This study was aimed at developing a prediction model for the diagnosis of EC in the general population. First, we obtained datasets GSE63678, GSE106191, and GSE115810 from the Gene Expression Omnibus (GEO) database, dataset GSE17025 from the GEO database, and the RNA sequence of EC from The Cancer Genome Atlas (TCGA) database to constitute the training, test, and validation groups, respectively. Subsequently, the 96 most significantly differentially expressed genes (DEGs) were identified and analyzed for function and pathway enrichment in the training group. Next, we acquired the disease-specific genes by random forest and established an artificial neural network for the diagnosis. Receiver operating characteristic (ROC) curves were utilized to identify the signature across the three groups. Finally, immune infiltration was analyzed to reveal tumor-immune microenvironment (TIME) alterations in EC. The top 96 DEGs (77 down-regulated and 19 up-regulated genes) were primarily enriched in the interleukin-17 signaling pathway, protein digestion and absorption, and transcriptional misregulation in cancer. Subsequently, 14 characterizing genes of EC were identified by random forest. In the training, test, and validation groups, the artificial neural network was constructed with high diagnostic accuracies of 0.882, 0.864, and 0.839, respectively, and areas under the ROC curve (AUCs) of 0.928, 0.921, and 0.782, respectively. Finally, resting and activated mast cells were found to have increased in TIME. We constructed an artificial diagnostic model with excellent reliability for EC and uncovered variations in the immunological ecosystem of EC through integrated bioinformatics approaches, which might be potential diagnostic targets for EC.

Keywords: endometrial carcinoma; GEO; TCGA; random forest; receiver operating characteristic curve

1. Introduction

Endometrial carcinoma (EC), a malignancy of the inner epithelial lining of the uterus, is a common neoplasm in women worldwide, with increasing rates of incidence and disease-associated mortality in recent years [1,2], seriously threatening women's physical and mental health. Most cases of early EC are cured by surgery alone or with adjuvant therapy. However, many cases of EC are diagnosed in the advanced stage at the first consultation and are associated with a poor prognosis. Although the survival rate of patients has increased, owing to molecular targeted therapy, no targeted gene mutations have been explored in advanced EC [3–5].

Currently, EC is diagnosed mainly based on clinical symptoms; physical findings; results of laboratory investigations, transvaginal ultrasound, pelvic ultrasonography, endometrial biopsy with hysteroscopy, and imaging (computed tomography, positron emission tomography/computed tomography, and magnetic resonance imaging); and some biomarkers (e.g., CA125 and HE4) [6–9]. The purpose of these investigations is to examine the endometrial cells, determine the disease extent, and detect the presence/absence of metastasis. Although these methods have good sensitivity for the diagnosis of EC, they have disadvantages, such as poor specificity (particularly transvaginal ultrasound), invasiveness, pain, and high cost. Therefore, improved examination techniques are urgently required, and target genes seem to be appropriate candidates.

Owing to advancements in computer technology and the introduction of sequencing technology, studies have promoted our understanding of cellular and genetic changes during oncogenesis and yielded more targeted and individualized treatment choices [10–12]. Machine learning, a component of artificial intelligence, using computer technology to simulate human intellect, can make predictions using mathematical algorithms after being trained with data. Deep learning, a branch of machine learning, focuses on making forecasts using a multilayer neural network algorithm and can expand model predictions exponentially with increased data volume and dimension, making it suitable for large-scale data analyses. Thus, deep learning can generate meaningful insights and discern relevant traits from genomic data. Genomic analyses have revealed novel biological targets for EC. The genetic bases of cancer progression and therapeutic response have been extensively studied, and the developments of next-generation sequencing and machine learning have yielded opportunities to systematically assess differentially expressed genes (DEGs) [11–13]. Moreover, large public databases, such as the Gene Expression Omnibus (GEO) and The Cancer Genome Atlas (TCGA), have provided abundant cancer genome sequencing data, which have improved our understanding of molecular changes in oncogenesis. However, due to the lack of multi-omics data, studies on the genomic analysis of EC focusing on gene expression or immune response are few. Since RNA sequencing of tumor tissues is usually performed to characterize gene expression and tumor immune microenvironment (TIME) cells, many datasets have estimated the abundance of DEGs and TIME cells in neoplastic tissue [13–16].

This study was aimed at identifying the signature genes in EC using machine learning, constructing a diagnostic model using an artificial neural network, and verifying the model in three EC cohorts. Finally, the changes in TIME during EC were confirmed.

2. Materials and Methods

2.1. Data Collection and Pre-Processing

Table 1 demonstrates the datasets utilized in this study. The gene expression datasets GSE63678, GSE106191, and GSE115810 were obtained from GEO (<https://www.ncbi.nlm.nih.gov/geo/>) (accessed on 3 March 2022), merged, and corrected for the batch effect to constitute the training group. The dataset GSE17025 from GEO and the gene expression of EC from TCGA (<https://www.cancer.gov/>) (accessed on 3 March 2022) were accessed to constitute the test and validation groups, respectively. Our study complied with the publication guidelines laid down by GEO and TCGA. No ethics committee approval was required.

Table 1. Composition of the datasets and component of patients enrolled in this study.

	Train Group		Test Group		Validation Group
	GSE106191	GSE115810	GSE63678	GSE17025	TCGA
Sample Count	97	27	35	103	583
Normal	64	3	5	12	35
Cancer	33	24	7	91	548
Enrollment	97	27	12	103	583

2.2. Exploration of DEGs and Functional Enrichment

The 96 DEGs across the EC and para-cancer samples in the training group were calculated using the R package “limma”, which employed the empirical Bayesian method and the moderated Wilcox test to assess differences in gene expression. Subsequently, heatmaps and volcanic maps were drawn using the R package with an absolute log₂ fold change ≥ 0.8 and an adjusted *p*-value < 0.05 . For the next functional analysis of the 96 DEGs in EC, the R package “clusterProfiler” was used to perform the Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGGs) enrichment analyses. The GO analysis mainly comprised the biological process, cellular component, and molecular function. For the functional enrichment analysis, statistical significance was set at *p* < 0.05 , and the R packages “enrichplot” and “ggplot2” were used.

2.3. Construction of Metascape and the Protein-Protein Interaction (PPI) Network

In addition, we also analyzed gene sets using the online toolkit WebGestalt (<http://www.webgestalt.org/>) (accessed on 3 March 2022); performed enrichment analyses using Metascape (<http://metascape.org/>) (accessed on 3 March 2022), Reactome, and WikipathwayCancer; and investigated a protein–protein interaction (PPI) network using the STRING (<https://cn.string-db.org/>) (accessed on 3 March 2022) database.

2.4. Selection of the Signature Genes and Construction of the Diagnostic Prediction Model

Random forest analyses were performed, and characteristic DEGs were selected based on the point at which the error of cross validation was the least. The setting seed was 123,456, and the ntree was 500. Subsequently, the characteristic genes were assigned a gene importance score, and those with a score > 0.9 were selected and visualized by the R packages “limma” and “pheatmap”. Next, we clustered the samples according to the expression of DEGs in the training group and found that the samples were divided into two clusters, similar to carcinoma and paraneoplastic samples.

Subsequently, we assigned scores to the specific DEGs to eliminate batch effects in samples. Up-regulated genes greater than the median value were scored 1, whereas the rest were scored 0; similarly, down-regulated genes lesser than the median value were scored 1, whereas the rest were scored 0. The artificial neural network model for the EC diagnosis was constructed from three types of layers: the input layer, with the scores of 14 genes; the hidden layers, with the scores and weights of genes; and the output layer, with the results for control and experimental samples. The R package “NeuralNetTools” was applied for the procedure with a seed of 12,345,678. Similarly, the selected DEGs and the constructed artificial neural network were applied to the test and validation groups. Unlike the other two groups, the control samples enrolled in the test group comprised tissues of other uterine pathologic types, whereas the samples in the experimental group comprised tissues of early EC. In addition, we constructed a receiver operating characteristic (ROC) curve using the R package “pROC” and assessed the area under the ROC curve (AUC) for the diagnostic model across the three cohorts.

2.5. Identification of TIME

In the analysis of immune cell infiltration, a total of 22 immune cells were identified by the CIBERSORT algorithm and screened using the R packages “e1071”, “preprocessCore”, and “CIBERSORT.R” at $p < 0.05$. The correlation between the immune cells was calculated using the R package “corrplot.” Moreover, the different distribution of immune cells between EC and normal tissues was measured and presented as a violin plot.

3. Results

3.1. DEGs and Functional Enrichment Analysis Results in EC

Based on the filer criteria, a total of 96 DEGs were found between EC and normal samples in the training group and analyzed. There were 19 up-regulated (e.g., MMP12 and CCL20) and 77 down-regulated (e.g., SFP4, OGN, OSR2, FOXL2, and IGFBP4) genes (Figure 1A,B). The top 10 GO terms revealed that the DEGs were mainly involved in collagen-containing extracellular matrix organization and signaling receptor activator activity (Figure 1C). KEGG terms demonstrated that the 96 DEGs were mainly involved in the interleukin-17 (IL-17) signaling pathway, protein digestion and absorption, and transcriptional misregulation in cancer (Figure 1D); thereby playing important roles in inflammatory and immune processes and the occurrence and development of tumors. All enrichment analysis results were closely related to TIME.

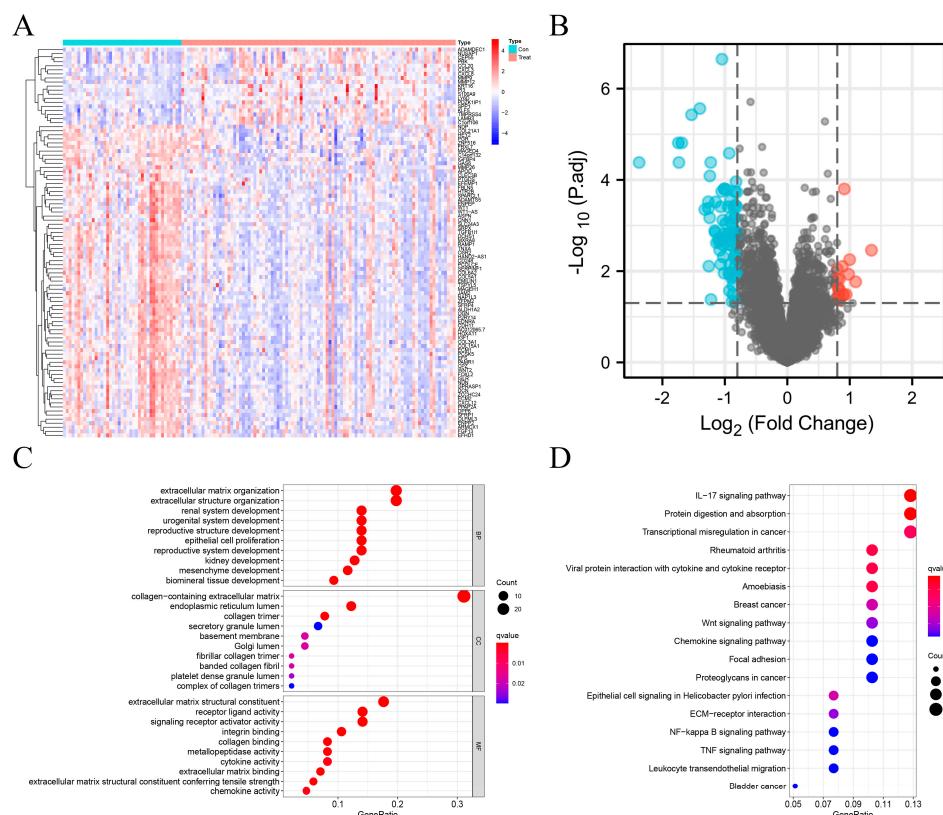


Figure 1. Identification of 96 DEGs in EC in the training group. **(A)** Heatmap of DEGs. The columns represent samples, and rows represent genes. The red color represents up-regulation, and the blue color represents down-regulation. $|\log_2\text{FC}| > 0.8$, $p\text{-value} < 0.05$. **(B)** Volcanic map of DEGs. The red, blue, and black colors represent up-regulated, down-regulated, and undifferentiated genes, respectively. $|\log_2\text{FC}| > 0.8$, $p\text{-value} < 0.05$. **(C)** Top 10 biological processes, cellular components, and molecular functions with the most significant p -value. **(D)** All KEGG enrichment results of DEGs.

3.2. Metascape and PPI Network Analysis Results

A network diagram was created based on Metascape analysis. Spots represented functions or pathways. Larger and connected points represented the presence of more similar genes between the functions or pathways. The NABA_CORE_MATRISOME gene set contained many genes encoding extracellular matrix organization and extracellular matrix-associated proteins activated in EC, while the NABA_MATRISOME_ASSOCIATED gene set contained many genes encoding vascular development, tissue morphogenesis, and growth regulation (Figure 2A). Figure 2B shows the top 50 function enrichments. Subsequently, the enrichment analyses of DisGeNET and PaGenBase revealed that the DEGs were primarily specialized in endometrial neoplasms and the uterus (Figure 2C,D), consistent with this study. During the pathogenesis of EC, epigenetic changes in pathogenic genes were mainly regulated by transcription factors EP300, RELA, JUN, SP1, NFKB1, ERG, HDAC1, CEBPA, FOS, and HIF1A (Figure 2E), which play important roles in inflammation, cell proliferation, transformation, differentiation, apoptosis, and immune response. In addition, the PPI network showed a relationship between different genes and proteins in the three sub-modules (Figure 2F). The NABA_CORE_MATRISOME sub-module included COL21A1, COL5A1, COL6A2, COL3A1, and COL15A1, which can identify the structural components of the extracellular matrix to provide tensile strength; the extracellular matrix organization sub-module included SPP1, IGFBP4, GAS6, MXRA8, and SPARCL1, which could enable proteins and/or the extracellular matrix; the NABA_MATRISOME_ASSOCIATED sub-module included P2RY14, CXCL8, CCL20, CXCL3, and CXCL12, which could enable protein binding and chemokine activity.

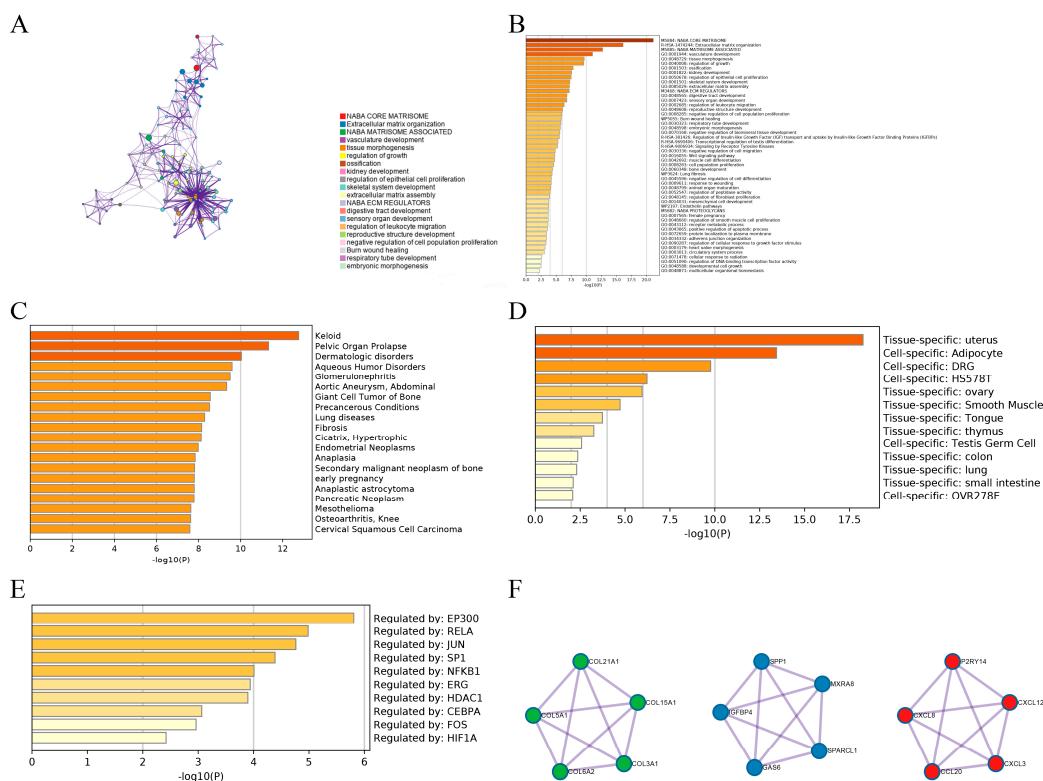


Figure 2. PPI network based on Metascape. (A) Network diagrams of the enrichment pathway and process of EC. (B) Bar plot of the enrichment pathway and process of EC. (C) Bar plot of enrichment on DisGeNET. (D) Bar plot of enrichment on PaGenBase. (E) Bar chart of enrichment on TRRUST. (F) Three sub-modules of PPI.

3.3. Exploration of Characteristic DEGs and Diagnostic Prediction Model of EC

We conducted a random forest analysis to identify the characteristic DEGs. The black line and horizontal and vertical axes represented the error value of the samples, number of trees, and cross-validation error, respectively (Figure 3A). Figure 3B shows the importance of genes. After re-validating DEGs, all 14 EC-signature DEGs with a score >0.9 were enrolled, including three up-regulated (MMP12, MMP9, and ADAMDEC1) and 11 down-regulated (OGN, FOXL2, IGFBP4, DCHS1, ENPP2, ALDH1A2, ADAMTS5, MXRA8, EFEMP1, EFS, and ENPEP genes (Figures 3C and 4). In the diagnostic prediction model, the control and experimental samples were aggregated, which signified that the expression of the pathogenic genes was distinguished between the normal and EC samples (Figure 3D). In addition, for the training, test, and validation groups, the AUCs were 0.928, 0.921, and 0.782, respectively, and the accuracies were 0.882, 0.864 and 0.839, respectively (Figure 5A–C, Table 2); implying that the EC diagnostic prediction model could be used as an independent diagnostic predictor of EC.

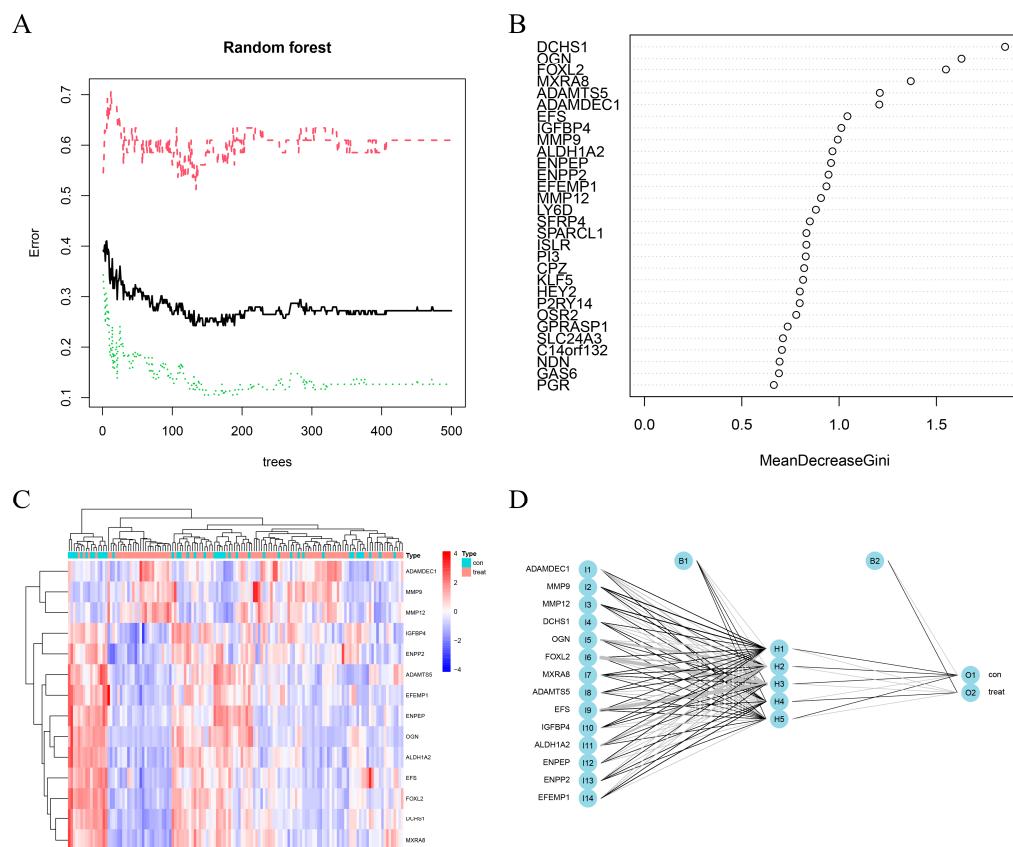


Figure 3. Selection of signature genes by machine learning and construction of a diagnostic prediction model by artificial neural network. (A) Construction of random forest. (B) Exploring signature genes of EC based on gene importance scores. (C) Heatmap of 14 characteristic DEGs. (D) Process of constructing artificial neural network.

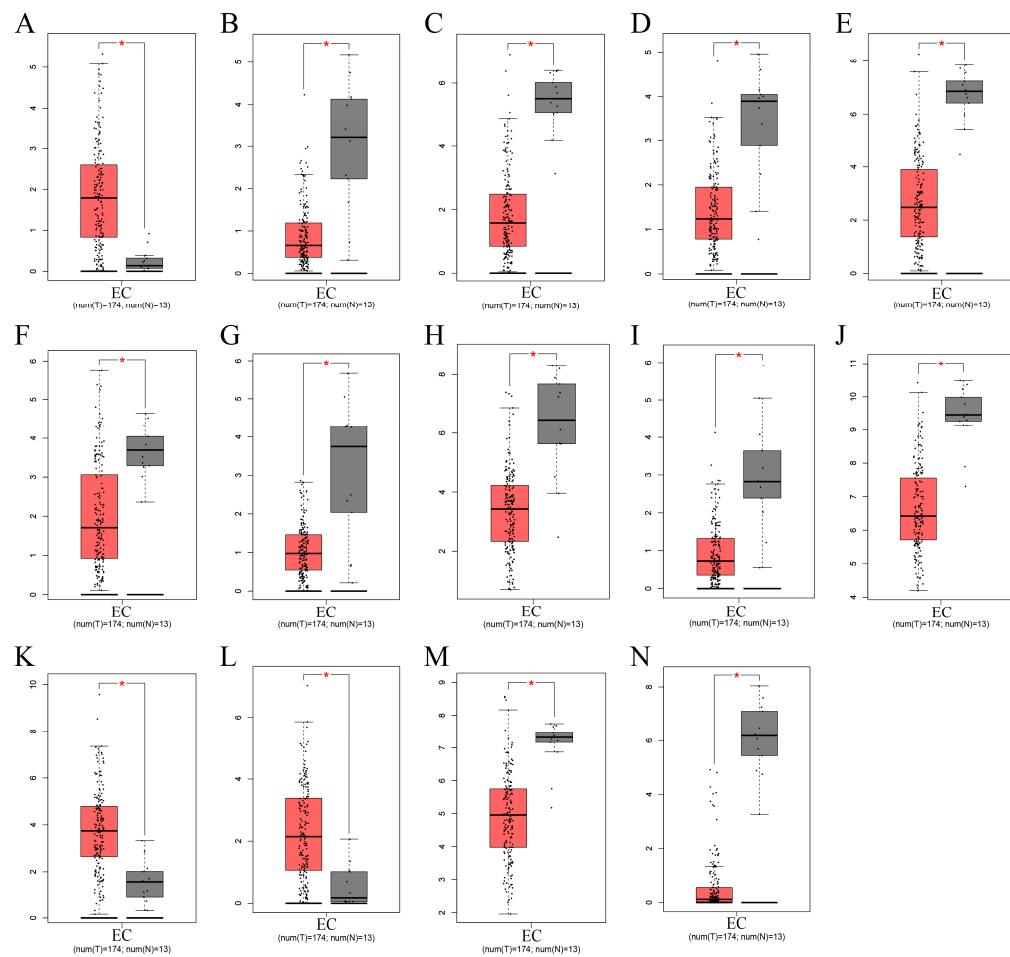


Figure 4. Box diagram of 14 characteristic genes in EC and healthy controls with p -value < 0.01 . (A–N) ADAMDEC1, ADAMTS5, ALDH1A2, DCHS1, EFEMP1, EFS, ENPEP, ENPP2, FOXL2, IGFBP4, MMP9, MMP12, MXRA8, and OGN. The red color represents EC, and the black color represents healthy controls. * means p -value < 0.01 .

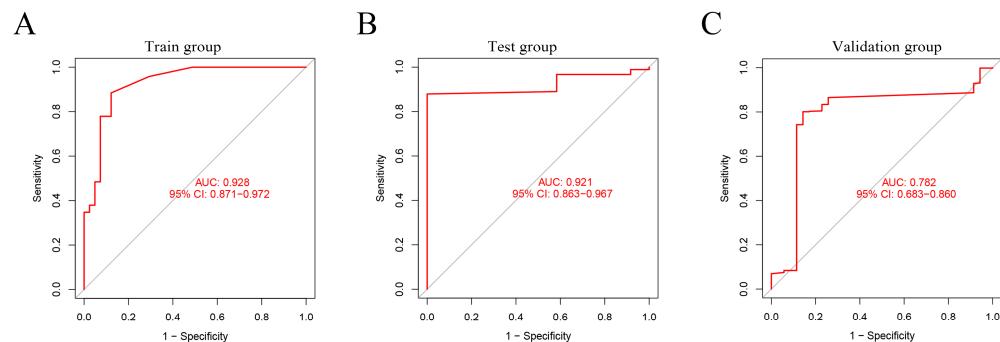


Figure 5. ROC curves of the three groups. (A) Training group. (B) Test group. (C) Validation group.

Table 2. Neural Diagnostic for the training, test and validation cohorts.

		Training Group		Test Group		Validation Group	
		Normal	Cancer	Normal	Cancer	Normal	Cancer
Prediction results	Normal	32	7	12	14	26	85
	Cancer	9	88	0	77	9	463
	Normal Accuracy	0.780		1.000		0.743	
	Cancer Accuracy	0.926		0.846		0.845	
Accuracy		0.882		0.864		0.839	

3.4. TIME of EC

Figure 6A shows the 22 categories of immunocytes in each sample. Resting and activated mast cells, neutrophils, macrophage M1s, activated NK cells, and eosinophils were relatively abundant in EC. Figure 6B shows the correlation in infiltration of immune cells. The greater the absolute value of the number, the stronger the correlation coefficients, with red and blue colors representing positive and negative correlations, respectively. Activated and resting mast cells showed a strong negative correlation, with a correlation coefficient of -0.54 . Activated mast cells and NK cells showed a negative correlation, with a correlation coefficient of -0.43 . The activated T cells CD4 and CD8 showed a strong positive correlation, with a correlation coefficient of 0.36 (Figure 6B). In summary, resting and activated mast cells, neutrophils, macrophage M1s, activated NK cells, and eosinophils in EC and normal samples were significantly different (Figure 6A–C); high expressions of activated mast cells, macrophage M1, and neutrophils and low expressions of resting mast cells, activated NK cells, and eosinophils were found in EC.

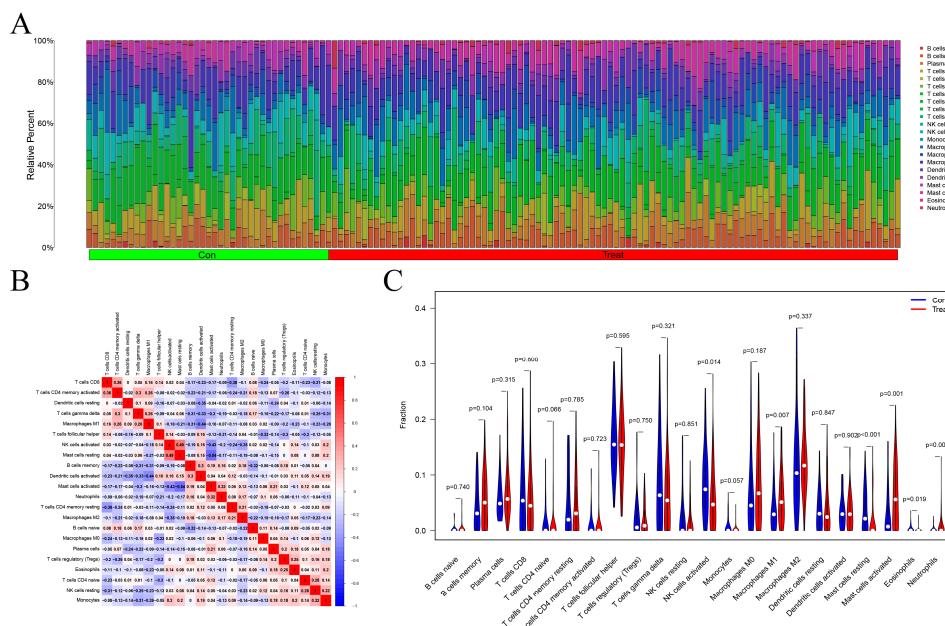


Figure 6. Tumor-immune microenvironment of EC. (A) Histogram of 22 types of immune cells in EC and healthy controls. (B) Correlation of immune cells in EC. (C) Violin image of immune cells.

4. Discussion

At present, EC is diagnosed mainly based on clinical symptoms, physical findings, results of laboratory investigations, and imaging examination. Endometrial biopsy under hysteroscopy seems to be the best method for the diagnosis of benign EC [17,18]. Fertility retention technology can effectively improve the quality of life of gynecological cancer patients, and has become the goal and hope for cancer survivors to live a better life [19]. Studies based on systems biology proteomics have highlighted the exact potential molecular mechanisms associated with SLN and EC grades [20,21]. The aim of

these investigations is to examine the endometrial cells, determine the disease extent, and detect the presence/absence of metastasis. Although the accuracy of the diagnosis and treatment of EC has made great progress in recent years, the molecular mechanism remains unknown. Abnormal gene expression and immune response in TIME play active roles in tumor occurrence, development, invasion, metastasis, and recurrence and are key considerations influencing tumor prognosis [16,22–24]. Endometrial biopsy under hysteroscopy seems to be the best method for the diagnosis of benign EC. In this study, we focused on transcriptional data from GEO and TCGA to identify the complex correlations of the signature genes for EC with a diagnosis and to build a diagnostic prediction model of EC, involving 14 signature genes by random forest and artificial neural network analyses, which distinguished patients with EC from the general population to guide diagnosis and treatment.

We obtained 96 DEGs, including 19 up-regulated and 77 down-regulated genes, and investigated sophisticated biological functions using GO and KEGG analyses in the training group. The outcome indicated that the DEGs were mainly enriched in extracellular matrix and structure organizations, and involved in the IL-17 signaling pathway, protein digestion and absorption, and transcriptional mis-regulation in TIME. These results indicated that changes in gene expression could be conducive to tumor remodeling and promote chronic inflammation, tumor progression, metastasis, and immune escape. We also obtained ten transcription factors, including EP300, RELA, JUN, SP1, NFKB1, ERG, HDAC1, CEBPA, FOS, and HIF1A, which regulated gene expression and played important roles in inflammation, cell proliferation, transformation, differentiation, apoptosis, and immune response [25–30]. In addition, the PPI network mainly showed a relationship between different genes and proteins among the three sub-modules. The NABA_CORE_MATRISOME sub-module comprised COL21A1, COL5A1, COL6A2, COL3A1, and COL15A1, which could identify structural components of the extracellular matrix to provide tensile strength. The extracellular matrix organization sub-module comprised SPP1, IGFBP4, GAS6, MXRA8, and SPARCL1, which could enable proteins and extracellular matrix. The NABA_MATRISOME_ASSOCIATED sub-module comprised P2RY14, CXCL8, CCL20, CXCL3, and CXCL12, which could enable protein binding and chemokine activity.

To obtain a good neural network model, we found 14 characteristic genes for EC by the machine learning method random forest. A diagnostic prediction model for EC was constructed using the artificial neural network, which may be widely applied to the formulation of diagnosis and treatment models for EC. In the model, expressions of MMP12, MMP9, and ADAMDEC1 were increased in EC, and those of OGN, FOXL2, IGFBP4, DCHS1, ENPP2, ALDH1A2, ADAMTS5, MXRA8, EFEMP1, EFS, and ENPEP were decreased in EC. MMP12 and MMP9 were related to cancer development, progression, and survival through various pathological processes and play essential roles in tumor invasion and metastasis [31–34]. Therefore, MMP12 knockdown inhibited proliferation and invasion of nasopharyngeal and lung cancers. Overexpression of ADAMDEC1 is correlated with tumor progression, inflammation, immunotherapeutic response, and a poor prognosis in many cancers [35–37]. Under-expressed OGN and EFS, compared to the normal samples, improved survival, reduced tumor recurrence, and reversed the epithelial to mesenchymal transition by inhibiting EGFR/AKT/Zeb-1 in tumors [38,39]. In a previous study, FOXL2 was considered for molecular diagnostic testing in ovarian adult granulosa cell and microcystic stromal cancers [40]. IGFBP-4 plays an important role in tumor growth regulation by inhibiting IGF actions [41]. Although these feature genes are widely expressed in tumors, according to previous reports, further research is required to clarify the gene function in the pathology of carcinoma, particularly EC. According to the traditional model, EC is divided into types 1 and 2, with certain classic mutations between the two types. Type 1 has mutations in PTEN, ARID1A, PIK3CA, and KRAS, while type 2 has mutations in TP53. Currently, EC is mainly diagnosed based on uterine curettage or biopsy findings. Some data suggest that the susceptibility of endometrial biopsy for EC is 52–94% [42–46].

The accuracy of differentiation of EC in other studies was slightly lower than our model (Table 2) [21,47,48]. Particularly, the test group comprised non-cancerous uterine pathologic types and early EC. The diagnostic rate of 100% in the non-cancerous group confirmed the efficacy of our diagnostic model for early EC in the test group. Thus, the model in the training and test groups showed a good effect, while that in the validation group showed an average effect. The 14 feature genes were key potential biomarkers of EC, but further studies are required to verify the results.

In addition, we also focused on TIME of EC and found that high expressions of activated mast cells, macrophage M1s, and neutrophils, and low expressions of resting mast cells, NK cells, and activated eosinophils played vital roles in EC. Multiple studies have documented that mast cells, neutrophils, macrophage M1s, NK cells, and eosinophils play a protective role during cancer progression, such as inflammatory responses, development of blood vessels, apoptosis, proliferation, invasion, and immune evasion [49–56].

However, this study has some limitations. First, the RNA sequencing data were only obtained from public databases. Second, although we validated the predictive performance of the EC diagnosis, further investigation is required for accurate validation. Further basic and clinical studies should be performed to validate the outcome and find a simpler, faster, and more economic approach.

5. Conclusions

In our study, we identified 14 genes involved in EC, verified them, based on GEO and TCGA, and established a robust diagnostic prediction model for EC through an artificial neural network, which was promising for the exploration of new diagnostic tools. The diagnostic model possessed excellent sensitivity and specificity, demonstrating the capability of diagnosing early EC. We also discovered that activated and resting mast cells were important and inversely correlated in EC. These results could serve as a basis for extensive cohorts in the future.

Author Contributions: All authors made important contributions to the study design, data acquisition, and data analysis; formal analysis, D.Z. and Z.W.; investigation, D.Z. and Z.W.; writing—original draft preparation, D.Z. and Z.Z.; writing—review and editing, D.Z. and Z.D.; visualization, M.W., T.Z. and J.Z.; supervision W.Z. and Y.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

Endometrial Carcinoma: EC; Gene Expression Omnibus: GEO; The Cancer Genome Atlas: TCGA; Differentially Expressed Genes: DEGs; Tumor Immune Microenvironment: TIME; Support Vector Machine: SVM; Protein–Protein Interaction: PPI; Gene Ontology: GO; Kyoto Encyclopedia of Genes and Genomes: KEGG Receiver Operating Characteristic: ROC; Area Under Curve: AUC.

References

1. Sung, H.; Ferlay, J.; Siegel, R.L.; Laversanne, M.; Soerjomataram, I.; Jemal, A.; Bray, F. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J. Clin.* **2021**, *71*, 209–249. [[CrossRef](#)] [[PubMed](#)]
2. Koh, W.J.; Abu-Rustum, N.R.; Bean, S.; Bradley, K.; Campos, S.M.; Cho, K.R.; Chon, H.S.; Chu, C.; Cohn, D.; Crispens, M.A.; et al. Uterine Neoplasms, Version 1.2018, NCCN Clinical Practice Guidelines in Oncology. *J. Natl. Compr. Cancer Netw.* **2018**, *16*, 170–199. [[CrossRef](#)] [[PubMed](#)]

3. Brooks, R.A.; Fleming, G.F.; Lastra, R.R.; Lee, N.K.; Moroney, J.W.; Son, C.H.; Tatebe, K.; Veneris, J.L. Current recommendations and recent progress in endometrial cancer. *CA Cancer J. Clin.* **2019**, *69*, 258–279. [[CrossRef](#)] [[PubMed](#)]
4. Bolivar, A.M.; Luthra, R.; Mehrotra, M.; Chen, W.; Barkoh, B.A.; Hu, P.; Zhang, W.; Broaddus, R.R. Targeted next-generation sequencing of endometrial cancer and matched circulating tumor DNA: Identification of plasma-based, tumor-associated mutations in early stage patients. *Mod. Pathol.* **2019**, *32*, 405–414. [[CrossRef](#)]
5. Bell, D.W.; Ellenson, L.H. Molecular Genetics of Endometrial Carcinoma. *Annu. Rev. Pathol.* **2019**, *14*, 339–367. [[CrossRef](#)]
6. McKenney, J.K.; Longacre, T.A. Low-grade endometrial adenocarcinoma: A diagnostic algorithm for distinguishing atypical endometrial hyperplasia and other benign (and malignant) mimics. *Adv. Anat. Pathol.* **2009**, *16*, 1–22. [[CrossRef](#)] [[PubMed](#)]
7. Gimpelson, R.J.; Rappold, H.O. A comparative study between panoramic hysteroscopy with directed biopsies and dilatation and curettage. A review of 276 cases. *Am. J. Obstet. Gynecol.* **1988**, *158*, 489–492. [[CrossRef](#)]
8. Antonsen, S.L.; Jensen, L.N.; Loft, A.; Berthelsen, A.K.; Costa, J.; Tabor, A.; Qvist, I.; Hansen, M.R.; Fisker, R.; Andersen, E.S.; et al. MRI, PET/CT and ultrasound in the preoperative staging of endometrial cancer—A multicenter prospective comparative study. *Gynecol. Oncol.* **2013**, *128*, 300–308. [[CrossRef](#)]
9. Duk, J.M.; Aalders, J.G.; Fleuren, G.J.; de Bruijn, H.W. CA 125: A useful marker in endometrial carcinoma. *Am. J. Obstet. Gynecol.* **1986**, *155*, 1097–1102. [[CrossRef](#)]
10. Sone, K.; Toyohara, Y.; Taguchi, A.; Miyamoto, Y.; Tanikawa, M.; Uchino-Mori, M.; Iriyama, T.; Tsuruga, T.; Osuga, Y. Application of artificial intelligence in gynecologic malignancies: A review. *J. Obstet. Gynaecol. Res.* **2021**, *47*, 2577–2585. [[CrossRef](#)]
11. Hamamoto, R. Application of Artificial Intelligence for Medical Research. *Biomolecules* **2021**, *11*, 90. [[CrossRef](#)] [[PubMed](#)]
12. Hamamoto, R.; Komatsu, M.; Takasawa, K.; Asada, K.; Kaneko, S. Epigenetics Analysis and Integrated Analysis of Multiomics Data, Including Epigenetic Data, Using Artificial Intelligence in the Era of Precision Medicine. *Biomolecules* **2019**, *10*, 62. [[CrossRef](#)] [[PubMed](#)]
13. Welford, S.M.; Gregg, J.; Chen, E.; Garrison, D.; Sorensen, P.H.; Denny, C.T.; Nelson, S.F. Detection of differentially expressed genes in primary tumor tissues using representational differences analysis coupled to microarray hybridization. *Nucleic Acids Res.* **1998**, *26*, 3059–3065. [[CrossRef](#)] [[PubMed](#)]
14. Albaradei, S.; Thafar, M.; Alsaedi, A.; Van Neste, C.; Gojobori, T.; Essack, M.; Gao, X. Machine learning and deep learning methods that use omics data for metastasis prediction. *Comput. Struct. Biotechnol. J.* **2021**, *19*, 5008–5018. [[CrossRef](#)]
15. Jiménez-Sánchez, D.; Ariz, M.; Chang, H.; Matias-Guiu, X.; de Andrea, C.E.; Ortiz-de-Solórzano, C. NaroNet: Discovery of tumor microenvironment elements from highly multiplexed images. *Med. Image Anal.* **2022**, *78*, 102384. [[CrossRef](#)]
16. Ruan, T.; Wan, J.; Song, Q.; Chen, P.; Li, X. Identification of a Novel Epithelial-Mesenchymal Transition-Related Gene Signature for Endometrial Carcinoma Prognosis. *Genes* **2022**, *13*, 216. [[CrossRef](#)]
17. Vitale, S.G.; Riemma, G.; Carugno, J.; Chiofalo, B.; Vilos, G.A.; Cianci, S.; Budak, M.S.; Lasmar, B.P.; Raffone, A.; Kahramanoglu, I. Hysteroscopy in the management of endometrial hyperplasia and cancer in reproductive aged women: New developments and current perspectives. *Transl. Cancer Res.* **2020**, *9*, 7767–7777. [[CrossRef](#)]
18. Príp, C.M.; Stentebjerg, M.; Bennetsen, M.H.; Petersen, L.K.; Bor, P. Risk of atypical hyperplasia and endometrial carcinoma after initial diagnosis of non-atypical endometrial hyperplasia: A long-term follow-up study. *PLoS ONE* **2022**, *17*, e0266339. [[CrossRef](#)]
19. La Rosa, V.L.; Garzon, S.; Gullo, G.; Fichera, M.; Sisti, G.; Gallo, P.; Rienna, G.; Schiattarella, A. Fertility preservation in women affected by gynaecological cancer: The importance of an integrated gynaecological and psychological approach. *Ecancermedicascience* **2020**, *14*, 1035. [[CrossRef](#)]
20. Aboulouard, S.; Wisztorski, M.; Duhamel, M.; Saudemont, P.; Cardon, T.; Narducci, F.; Lemaire, A.S.; Kobeissy, F.; Leblanc, E.; Fournier, I.; et al. In-depth proteomics analysis of sentinel lymph nodes from individuals with endometrial cancer. *Cell Rep. Med.* **2021**, *2*, 100318. [[CrossRef](#)]
21. Della Corte, L.; Giampaolino, P.; Mercurio, A.; Rienna, G.; Schiattarella, A.; De Franciscis, P.; Bifulco, G. Sentinel lymph node biopsy in endometrial cancer: State of the art. *Transl. Cancer Res.* **2020**, *9*, 7725–7733. [[CrossRef](#)] [[PubMed](#)]
22. Rousset-Rouviere, S.; Rochigneux, P.; Chrétien, A.S.; Fattori, S.; Gorvel, L.; Provansal, M.; Lambaudie, E.; Olive, D.; Sabatier, R. Endometrial Carcinoma: Immune Microenvironment and Emerging Treatments in Immuno-Oncology. *Biomedicines* **2021**, *9*, 632. [[CrossRef](#)] [[PubMed](#)]
23. Zheng, M.; Hu, Y.; Gou, R.; Li, S.; Nie, X.; Li, X.; Lin, B. Development of a seven-gene tumor immune microenvironment prognostic signature for high-risk grade III endometrial cancer. *Mol. Ther. Oncolytics* **2021**, *22*, 294–306. [[CrossRef](#)]
24. Chen, Y.; Lee, K.; Liang, Y.; Qin, S.; Zhu, Y.; Liu, J.; Yao, S. A Cholesterol Homeostasis-Related Gene Signature Predicts Prognosis of Endometrial Cancer and Correlates With Immune Infiltration. *Front. Genet.* **2021**, *12*, 763537. [[CrossRef](#)] [[PubMed](#)]
25. Ahn, S.H.; Edwards, A.K.; Singh, S.S.; Young, S.L.; Lessey, B.A.; Tayade, C. IL-17A Contributes to the Pathogenesis of Endometriosis by Triggering Proinflammatory Cytokines and Angiogenic Growth Factors. *J. Immunol.* **2015**, *195*, 2591–2600. [[CrossRef](#)]
26. Miossec, P.; Korn, T.; Kuchroo, V.K. Interleukin-17 and type 17 helper T cells. *N. Engl. J. Med.* **2009**, *361*, 888–898. [[CrossRef](#)]
27. Cornelius, D.C.; Lamarca, B. TH17- and IL-17- mediated autoantibodies and placental oxidative stress play a role in the pathophysiology of pre-eclampsia. *Minerva Ginecol.* **2014**, *66*, 243–249.
28. Liu, L.; Chen, F.; Xiu, A.; Du, B.; Ai, H.; Xie, W. Identification of Key Candidate Genes and Pathways in Endometrial Cancer by Integrated Bioinformatical Analysis. *Asian Pac. J. Cancer Prev.* **2018**, *19*, 969–975.
29. Gorczynski, R.M. IL-17 Signaling in the Tumor Microenvironment. *Adv. Exp. Med. Biol.* **2020**, *1240*, 47–58.

30. Lee, T.I.; Young, R.A. Transcriptional regulation and its misregulation in disease. *Cell* **2013**, *152*, 1237–1251. [CrossRef]
31. Gialeli, C.; Theocharis, A.D.; Karamanos, N.K. Roles of matrix metalloproteinases in cancer progression and their pharmacological targeting. *FEBS J.* **2011**, *278*, 16–27. [CrossRef] [PubMed]
32. Zheng, J.; Chu, D.; Wang, D.; Zhu, Y.; Zhang, X.; Ji, G.; Zhao, H.; Wu, G.; Du, J.; Zhao, Q. Matrix metalloproteinase-12 is associated with overall survival in Chinese patients with gastric cancer. *J. Surg. Oncol.* **2013**, *107*, 746–751. [CrossRef]
33. Brun, J.L.; Cortez, A.; Lesieur, B.; Uzan, S.; Rouzier, R.; Daraï, E. Expression of MMP-2, -7, -9, MT1-MMP and TIMP-1 and -2 has no prognostic relevance in patients with advanced epithelial ovarian cancer. *Oncol. Rep.* **2012**, *27*, 1049–1057. [CrossRef] [PubMed]
34. Wang, X.; Chen, T. CUL4A regulates endometrial cancer cell proliferation, invasion and migration by interacting with CSN6. *Mol. Med. Rep.* **2021**, *23*, 23. [CrossRef] [PubMed]
35. Liu, X.; Huang, H.; Li, X.; Zheng, X.; Zhou, C.; Xue, B.; He, J.; Zhang, Y.; Liu, L. Knockdown of ADAMDEC1 inhibits the progression of glioma in vitro. *Histol. Histopathol.* **2020**, *35*, 997–1005. [PubMed]
36. Zhu, W.; Shi, L.; Gong, Y.; Zhuo, L.; Wang, S.; Chen, S.; Zhang, B.; Ke, B. Upregulation of ADAMDEC1 correlates with tumor progression and predicts poor prognosis in non-small cell lung cancer (NSCLC) via the PI3K/AKT pathway. *Thorac. Cancer* **2022**, *13*, 1027–1039. [CrossRef]
37. Ahn, S.B.; Sharma, S.; Mohamedali, A.; Mahboob, S.; Redmond, W.J.; Pascovici, D.; Wu, J.X.; Zaw, T.; Adhikari, S.; Vaibhav, V.; et al. Potential early clinical stage colorectal cancer diagnosis using a proteomics blood test panel. *Clin. Proteom.* **2019**, *16*, 34. [CrossRef]
38. Lomnytska, M.I.; Becker, S.; Hellman, K.; Hellström, A.C.; Souchelnytskyi, S.; Mints, M.; Hellman, U.; Andersson, S.; Auer, G. Diagnostic protein marker patterns in squamous cervical cancer. *Proteom. Clin. Appl.* **2010**, *4*, 17–31. [CrossRef]
39. Hu, X.; Li, Y.Q.; Li, Q.G.; Ma, Y.L.; Peng, J.J.; Cai, S.J. Osteoglycin (OGN) reverses epithelial to mesenchymal transition and invasiveness in colorectal cancer via EGFR/Akt pathway. *J. Exp. Clin. Cancer Res. CR* **2018**, *37*, 41. [CrossRef]
40. Rabban, J.T.; Karnezis, A.N.; Devine, W.P. Practical roles for molecular diagnostic testing in ovarian adult granulosa cell tumour, Sertoli-Leydig cell tumour, microcystic stromal tumour and their mimics. *Histopathology* **2020**, *76*, 11–24. [CrossRef]
41. Baxter, R.C. IGF binding proteins in cancer: Mechanistic and clinical insights. *Nat. Rev. Cancer* **2014**, *14*, 329–341. [CrossRef] [PubMed]
42. Long, S. Endometrial Biopsy: Indications and Technique. *Primary care* **2021**, *48*, 555–567. [CrossRef] [PubMed]
43. Reijnen, C.; Visser, N.C.M.; Bulten, J.; Massuger, L.; van der Putten, L.J.M.; Pijnenborg, J.M.A. Diagnostic accuracy of endometrial biopsy in relation to the amount of tissue. *J. Clin. Pathol.* **2017**, *70*, 941–946. [CrossRef] [PubMed]
44. Kunaviktikul, K.; Suprasert, P.; Khunamornpong, S.; Settakorn, J.; Natpratan, A. Accuracy of the Wallach Endocell endometrial cell sampler in diagnosing endometrial carcinoma and hyperplasia. *J. Obstet. Gynaecol. Res.* **2011**, *37*, 483–488. [CrossRef] [PubMed]
45. Guido, R.S.; Kanbour-Shakir, A.; Rulin, M.C.; Christopherson, W.A. Pipelle endometrial sampling. Sensitivity in the detection of endometrial cancer. *J. Reprod. Med.* **1995**, *40*, 553–555.
46. Laban, M.; Nassar, S.; Elsayed, J.; Hassanin, A.S. Correlation between pre-operative diagnosis and final pathological diagnosis of endometrial malignancies; impact on primary surgical treatment. *Eur. J. Obstet. Gynecol. Reprod. Biol.* **2021**, *263*, 100–105. [CrossRef]
47. Della Corte, L.; Vitale, S.G.; Foreste, V.; Rienna, G.; Ferrari, F.; Noventa, M.; Liberto, A.; De Franciscis, P.; Tesarik, J. Novel diagnostic approaches to intrauterine neoplasm in fertile age: Sonography and hysteroscopy. *Off. J. Soc. Minim. Invasive Ther.* **2021**, *30*, 288–295. [CrossRef]
48. Heremans, R.; Van den Bosch, T.; Valentin, L.; Wynants, L.; Pascual, M.A.; Fruscio, R.; Testa, A.C.; Buonomo, F.; Guerriero, S.; Epstein, E.; et al. Ultrasound features of endometrial pathology in women without abnormal uterine bleeding: Results from the International Endometrial Tumor Analysis Study (IETA3). *Ultrasound Obstet. Gynecol.* **2022**. [CrossRef]
49. Johansson, A.; Rudolfsson, S.; Hammarsten, P.; Halin, S.; Pietras, K.; Jones, J.; Stattin, P.; Egevad, L.; Granfors, T.; Wikström, P.; et al. Mast cells are novel independent prognostic markers in prostate cancer and represent a target for therapy. *Am. J. Pathol.* **2010**, *177*, 1031–1041. [CrossRef]
50. Sinnamon, M.J.; Carter, K.J.; Sims, L.P.; Lafleur, B.; Fingleton, B.; Matrisian, L.M. A protective role of mast cells in intestinal tumorigenesis. *Carcinogenesis* **2008**, *29*, 880–886. [CrossRef]
51. Fleischmann, A.; Schlomm, T.; Köllermann, J.; Sekulic, N.; Huland, H.; Mirlacher, M.; Sauter, G.; Simon, R.; Erbersdobler, A. Immunological microenvironment in prostate cancer: High mast cell densities are associated with favorable tumor characteristics and good prognosis. *Prostate* **2009**, *69*, 976–981. [CrossRef] [PubMed]
52. Coffelt, S.B.; Wellenstein, M.D.; de Visser, K.E. Neutrophils in cancer: Neutral no more. *Nat. Rev. Cancer* **2016**, *16*, 431–446. [CrossRef] [PubMed]
53. Shojaei, F.; Singh, M.; Thompson, J.D.; Ferrara, N. Role of Bv8 in neutrophil-dependent angiogenesis in a transgenic model of cancer progression. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 2640–2645. [CrossRef] [PubMed]
54. Spiegel, A.; Brooks, M.W.; Houshyar, S.; Reinhardt, F.; Ardolino, M.; Fessler, E.; Chen, M.B.; Krall, J.A.; DeCock, J.; Zervant-tonakis, I.K.; et al. Neutrophils Suppress Intraluminal NK Cell-Mediated Tumor Cell Clearance and Enhance Extravasation of Disseminated Carcinoma Cells. *Cancer Discov.* **2016**, *6*, 630–649. [CrossRef]

55. Boutilier, A.J.; Elsawa, S.F. Macrophage Polarization States in the Tumor Microenvironment. *Int. J. Mol. Sci.* **2021**, *22*, 6995. [[CrossRef](#)]
56. Jhunjhunwala, S.; Hammer, C.; Delamarre, L. Antigen presentation in cancer: Insights into tumour immunogenicity and immune evasion. *Nat. Rev. Cancer* **2021**, *21*, 298–312. [[CrossRef](#)]