

RESEARCH

Open Access



Expression of EMT-related genes in lymph node metastasis in endometrial cancer: a TCGA-based study

He Li¹, Junzhu Wang², Liwei Li¹, Luyang Zhao¹ and Zhiqi Wang^{1*}

Abstract

Background Endometrial cancer (EC) with metastasis in pelvic/para-aortic lymph nodes suggests an unsatisfactory prognosis. Nevertheless, there is still rare literature focusing on the role of epithelial-mesenchymal transition (EMT) in lymph node metastasis (LNM) in EC.

Methods Transcriptional data were derived from the TCGA database. Patients with stage IA–IIIC2 EC were included, constituting the LN-positive and LN-negative groups. To evaluate the extent of EMT, an EMT signature composed of 315 genes was adopted. The EMT-related genes (ERGs) were obtained from the dbEMT2 database, and the differentially expressed ERGs (DEERGs) between these two groups were screened. On the basis of DEERGs, pathway analysis was carried out. We eventually adopted the logistic regression model to build an ERG-based gene signature with predictive value for LNM in EC.

Results A total of 498 patients were included, with 75 in the LN-positive group. Median EMT score of tumor tissues from LN-negative group was -0.369 , while that from the LN-positive group was -0.296 ($P < 0.001$), which clearly exhibited a more mesenchymal phenotype for LNM cases on the EMT continuum. By comparing expression profiles, 266 genes were identified as DEERGs, in which 184 were upregulated and 82 were downregulated. In pathway analysis, various EMT-related pathways were enriched. DEERGs shared between molecular subtypes were comparatively few. The ROC curve and logistic regression analysis screened 7 genes with the best performance to distinguish between the LN-positive and LN-negative group, i.e., *CIRBP*, *DDR1*, *F2RL2*, *HOXA10*, *PPARGC1A*, *SEMA3E*, and *TGFB1*. A logistic regression model including the 7-gene-based risk score, age, grade, myometrial invasion, and histological subtype was built, with an AUC of 0.850 and a favorable calibration ($P = 0.074$). In the validation dataset composed of 83 EC patients, the model exhibited a satisfactory predictive value and was well-calibrated ($P = 0.42$).

Conclusion The EMT status and expression of ERGs varied in LNM and non-LNM EC tissues, involving multiple EMT-related signaling pathways. Aside from that, the distribution of DEERGs differed among molecular subtypes. An ERG-based gene signature including 7 DEERGs exhibited a desirable predictive value for LNM in EC, which required further validation based upon clinical specimens in the future.

Keywords Endometrial cancer, Lymph node metastasis, Epithelial-mesenchymal transition, The cancer genome atlas, Pathway, Gene signature, Logistic regression model, Nomogram

*Correspondence:

Zhiqi Wang
wangzqnet@sina.com

¹ Department of Obstetrics and Gynecology, Peking University People's Hospital, No. 11 Xizhimen South Street, Xicheng District, Beijing 100044, China

² The Big Data and Public Policy Laboratory, School of Government, Peking University, Beijing, China



Background

In 2022, it is expected to be an incidence of 84,520 of uterine corpus cancer in China and 17,543 new deaths, where endometrial cancer (EC) accounts for the vast majority [1]. EC with metastasis in pelvic/para-aortic lymph nodes (LNs), even with micrometastasis, is associated with at least the International Federation of Gynecology and Obstetrics (FIGO) stage IIIC1/2 and suggests a less satisfactory prognosis [2, 3]. LN metastasis (LNM) of cancer is a complex process including lymphangiogenesis and activation of epithelial-mesenchymal transition (EMT) to initiate metastasis and survival in the LN [4–12]. A substantial proportion of studies on EMT in EC have been published [13–17], but literature fixing attention on the role of EMT in LNM remains rare to be reported [18]. As persuasively demonstrated by relevant studies, for primary prevention, adherence to cancer prevention guidelines was negatively associated with EC risk [19, 20]. Nonetheless, for tertiary prevention of EC, the prediction and treatment of LNM remain unsatisfactory. The mechanism-level description of this process will be advantageous for us to facilitate the management of the lymphatic spread of EC.

For this purpose, we presented a retrospective study on the basis of The Cancer Genome Atlas (TCGA) to discuss the role of EMT in LNM. Apart from that, this research not only probed deep into differentially expressed EMT-related genes (DEERGs) and pathways associated with them, but also ultimately provided an ERG-based gene signature with predictive value for LNM in EC.

Materials and methods

We extracted mRNA expression profiles from the Uterine Corpus Endometrial Carcinoma (TCGA, PanCancer Atlas) database through cBioPortal [21, 22]. Especially, tumor grade and microsatellite status were obtained from Xena [23]. Cases with a non-endometrioid histology were categorized into G3. Patients with FIGO 2009 stage IA–IIIC2 EC were included in this research, and stage IIIC1/2 were considered as stage IA–IIIB with pelvic/para-aortic LNM, constituting two groups (LN-positive and LN-negative). There were cases annotated by FIGO 1988 staging system, and patients with FIGO 1988 stage IIIA were excluded from this study, since peritoneal cytology was contained in this stage, which was absent in FIGO 2009 staging system.

To quantitatively locate a tumor tissue on the EMT continuum, a validated transcriptome-based scoring system composed of 315 genes (epithelial: 145, mesenchymal: 170) was adopted. An EMT score was calculated by the maximum vertical distance between the empirical distribution function (ECDF) curves of the

mesenchymal and the epithelial gene set by a two-sample Kolmogorov–Smirnov test (i.e., the KS method). The EMT score of a tumor with a higher expression of mesenchymal signatures would be defined as positive, and negative for epithelial ones. Hence, an EMT score in the KS method had a range of $[-1, 1]$ [24, 25] (Supplementary Fig. 1).

With regard to differential expression analysis, 1027 ERGs were obtained from an EMT gene database, “dbEMT2” [26], and DEERGs between the LN-positive and LN-negative groups were screened by the ‘limma’ package [27]. A false discovery rate (FDR) < 0.05 and $|\log_{2}FC| > 1$ were defined as the significance threshold. On the basis of DEERGs, Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analyses were carried out. The STRING database and Cytoscape were adopted to construct a protein–protein interaction (PPI) network to visualize the relationships among DEERGs, where 0.4 was defined as the minimum required interaction score, and genes with node degree > 15 were considered as hub genes [28–30].

For the sake of effectively screening candidate genes to construct a risk score for LNM, we adopted the “pROC” package to evaluate their predictive value [31]. An area under the curve (AUC) > 0.7 was considered acceptable for further evaluation. The logistic regression model was built to evaluate the predictive value of the genes screened, on the basis of which a nomogram was plotted. According to the probability predicted by the model, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) were reported. The consistency between the observed incidence rate and the predicted probability of the nomogram was evaluated by adopting a calibration plot with 2000 bootstrap replications. Meanwhile, the Hosmer and Lemeshow tests were used. We also introduced an EC database composed of 95 patients from the Clinical Proteomic Tumor Analysis Consortium (CPTAC) for external validation [32].

All the analyses were conducted by means of R software (version 4.1.1) [33]. Apart from the statistical methods above, the Wilcoxon test and chi-square test were adopted to compare continuous and categorical variables, respectively. A $P < 0.05$ was considered statistically significant.

Results

A total of 498 patients diagnosed between 1995 and 2013 were included in this study (Table 1), with 15.1% of patients exhibiting LNM. Tumor grade, histological subtype, myometrial invasion (MI), and molecular subtype were significantly associated with LNM ($P < 0.001$).

Table 1 Summary of the clinicopathological features of the patients stratified by LN status

Variables	LN-negative (n = 423)	LN-positive (n = 75)
Age (years), mean ± SD	63.8 ± 11.1	63.3 ± 10.4
Race, n (%)		
White	294 (69.5%)	45 (60.0%)
Black or African American	78 (18.4%)	21 (28.0%)
Asian	17 (4.0%)	2 (2.7%)
Others	8 (1.9%)	3 (4.0%)
Tumor grade, n (%)		
Grade1	95 (22.5%)	2 (2.7%)
Grade2	109 (25.8%)	8 (10.7%)
Grade3	219 (51.8%)	65 (86.7%)
Histological subtype, n (%)		
Endometrioid	350 (82.7%)	35 (46.7%)
Non-endometrioid	73 (17.3%)	40 (53.3%)
MI, n (%)		
< 1/2	285 (67.4%)	26 (34.7%)
≥ 1/2	138 (32.6%)	49 (65.3%)
Molecular subtype, n (%)		
POLE	39 (9.2%)	8 (10.7%)
MSI-H	134 (31.7%)	14 (18.7%)
CNL	139 (32.9%)	10 (13.3%)
CNH	103 (24.3%)	42 (56.0%)
Missing	8 (1.9%)	1 (1.3%)

Abbreviations: LN Lymph node, SD Standard deviation, EEC Endometrioid endometrial cancer, MI Myometrial invasion, MSI-H Microsatellite instability-high, CNL Copy-number low, CNH Copy-number high

EMT scores stratified by LN status

A total of 314 ERGs (144 epithelial and 170 mesenchymal) were included to evaluate the EMT status, with a pseudogene *OR7E14P* excluded on account of its absence in the TCGA database.

Coinciding with the epithelial origin of EC, general expressions of ERGs revealed a more epithelial phenotype [median -0.361, IQR (-0.449, -0.264)] in the whole cohort. The median EMT score of the LN-negative group was -0.369 [IQR (-0.455, -0.282)], while that of the LN-positive group was -0.296 [IQR (-0.449, -0.264)], with a $P < 0.001$ by Wilcoxon test, which presented a more mesenchymal phenotype for tumors with LNM on the EMT continuum (Fig. 1). Further subgroup analysis revealed that this phenotype difference was more significant in G3, $MI \geq 1/2$, and *POLE* cases (Supplementary Fig. 2).

Differentially expressed genes and pathway analysis

By comparing mRNA expression profiles of tumor tissues from 423 LN-negative cases and 75 LN-positive cases,

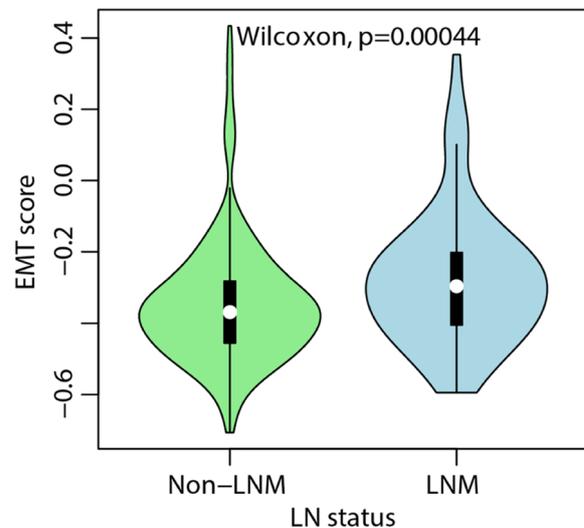


Fig. 1 EC tissues from LNM cases exhibited a more mesenchymal phenotype compared with their non-LNM counterparts

266 genes were identified as DEERGs, in which 184 were upregulated and 82 were downregulated. Expressions of ERGs were exhibited in Supplementary Fig. 3 and Supplementary Table 1.

GO enrichment terms on the basis of DEERGs were exhibited in Fig. 2. In the biological process (BP) group, various EMT-related processes were enriched (Supplementary Table 2), e.g., mesenchyme development ($q = 1.49 \times 10^{-15}$), mesenchymal cell differentiation ($q = 3.02 \times 10^{-14}$), epithelial to mesenchymal transition ($q = 1.97 \times 10^{-12}$), cellular response to transforming growth factor beta stimulus ($q = 4.14 \times 10^{-12}$), response to transforming growth factor beta ($q = 6.37 \times 10^{-12}$), exhibiting a propensity for the enrichment of TGF- β signaling pathway. A directed acyclic graph (DAG) displaying the correlation between the BPs enriched was as Supplementary Fig. 4.

KEGG pathway analysis exhibited a significant enrichment in EMT-related pathways, e.g., Notch signaling pathway, JAK-STAT signaling pathway, HIF-1 signaling pathway, TGF- β signaling pathway, and Wnt signaling pathway, most members of which exhibited an upregulation in LNM cases (Fig. 3, Supplementary Table 3).

The PPI network of DEERGs based upon the STRING database was presented in Supplementary Fig. 5. The top 5 hub genes calculated by Edge Percolated Component (EPC) method were *MYC*, *ESR1*, *KRAS*, *MAPK3*, and *HDAC1*. We also summarized hub genes by each method in Supplementary Table 4.

As is exhibited in Table 1, molecular subtypes of 96.8% of patients were available, and CNH cases were more likely to develop LNM than the others ($\chi^2 = 30.5$,

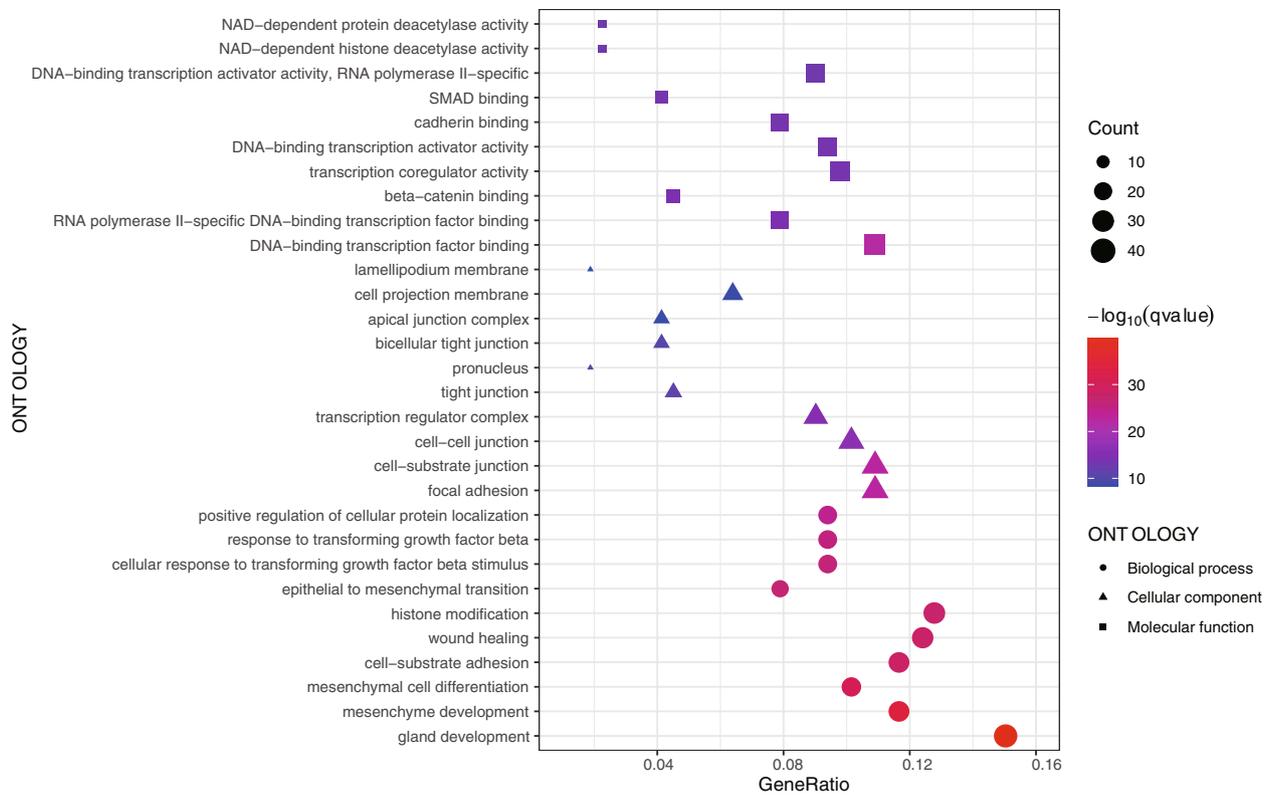


Fig. 2 Various EMT-related pathways were enriched, including the TGF- β signaling pathway. EMT, epithelial-mesenchymal transition; TGF- β , transforming growth factor-beta

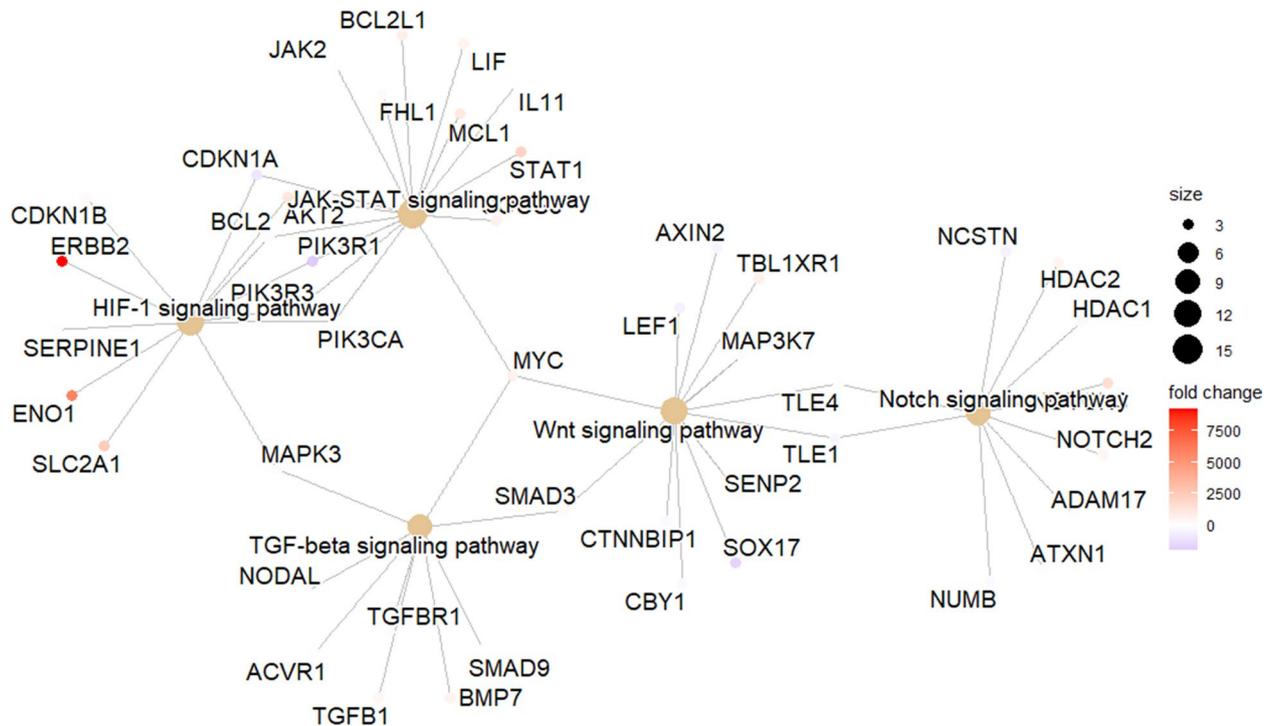


Fig. 3 Several EMT-related pathways enriched by KEGG. EMT, epithelial-mesenchymal transition; KEGG, Kyoto Encyclopedia of Genes and Genomes

$P < 0.001$). Therefore, we investigated DEERGs associated with LNM in cases with molecular subtypes ($n = 491$, Supplementary Table 5). A Venn graph displaying DEERGs shared between subtypes was depicted in Fig. 4, where DEERGs differed tremendously between subtypes. No DEERGs were shared between all subtypes, and only *HOXB7*, a gene encoding an activator of the TGF- β pathway, was shared by CNL, MSI-H, and *POLE*.

Identification of a gene signature effective to predict LNM

On the basis of receiver operating characteristic (ROC) curves adopted to discriminate between LN-positive and LN-negative groups, no DEERG exhibited an AUC > 0.7 (Supplementary Table 6). Consequently, a logistics regression model was constructed to screen candidate genes by including DEERG expression, age, race, grade, MI, and histological subtype, ROC curves from which were employed to screen target genes. As a result, the median AUC was 0.802 (IQR 0.801–0.805), which displays a comparatively high predictive value of the new models. To refine the candidate genes, genes from models with an AUC > 0.7 and $P < 0.05$ were selected for further analysis. Subsequently, 29 genes were selected (Supplementary Table 7), which were included in a further logistic regression model, and ultimately 7 genes

with the best performance were selected: *CIRBP*, *DDR1*, *F2RL2*, *HOXA10*, *PPARGC1A*, *SEMA3E*, *TGFB1* (Supplementary Table 8). An EMT-based risk score (ERS) was calculated as the logit from the logistic regression as follows: $1.845 + (0.323 \times CIRBP - 0.096 \times DDR1 - 1.381 \times F2RL2 - 0.190 \times HOXA10 - 2.348 \times PPARGC1A + 1.137 \times SEMA3E - 0.569 \times TGFB1) \times 10^{-3}$. Then we constructed a new logistic model to predict LNM with ERS, race, age, grade, MI, and histological subtype. A further stepwise algorithm filtered out race, eventually reaching an AUC of 0.850 (Fig. 5A). The cut-off value of the probability predicted by the logistic model is 0.176, with a sensitivity of 0.773, a specificity of 0.797, a PPV of 0.403, and an NPV of 0.952.

A nomogram based upon the above model was presented in Fig. 6, with ERS having the highest weight in the model compared with traditional clinicopathological parameters. A Hosmer–Lemeshow test exhibited a desirable calibration for this model ($\chi^2 = 14.3$, $P = 0.074$), with a calibration curve in Supplementary Fig. 6.

In an effort to validate the gene signature externally, we extracted mRNA expression profile of EC tissue from the CPTAC program, from which 83 cases identified as stage IA-IIIC2 were included. A similar logistic regression model was built (Supplementary Table 8), with an AUC of 0.908 (Fig. 5B) and an ideal calibration ($\chi^2 = 3.9$, $P = 0.42$). Aside from that, proteome from CPTAC exhibited a significantly positive correlation between proteins and mRNA in members of the 7-gene signature (Supplementary Fig. 7). Nevertheless, owing to the absence of *F2RL2* and *PPARGC1A* proteins, the value of a protein-based ERS was not assessed in this study.

Discussion

LNM had an incidence of 18.1% in EC, which was quite difficult to predict in clinical practice [3]. In our institution, EC patients with $MI \leq 1/2$, $G1/2$, and endometrioid histology, with no further evidence of invasion/metastasis, used to undergo a surgery without pelvic and para-aortic lymphadenectomy, all the others would receive lymphadenectomy, with sentinel lymphadenectomy (SLND) optional for patients of relatively low risk. Nevertheless, non-therapeutic systemic lymphadenectomy still existed in most cases (71.4%, unpublished data), which gives rise to an augment in the complication morbidity, e.g., blood loss, lymphatic cyst, and lower limb edema [34]. Meanwhile, variables employed to determine lymphadenectomy were mostly those significantly correlated with LNM. The molecular subtype has been proven to be significantly correlated with LNM [3], but has not been extensively adopted in the decision of lymphadenectomy yet. At the mechanism level, LNM is a complex biological process where EMT plays a central role in the invasion of

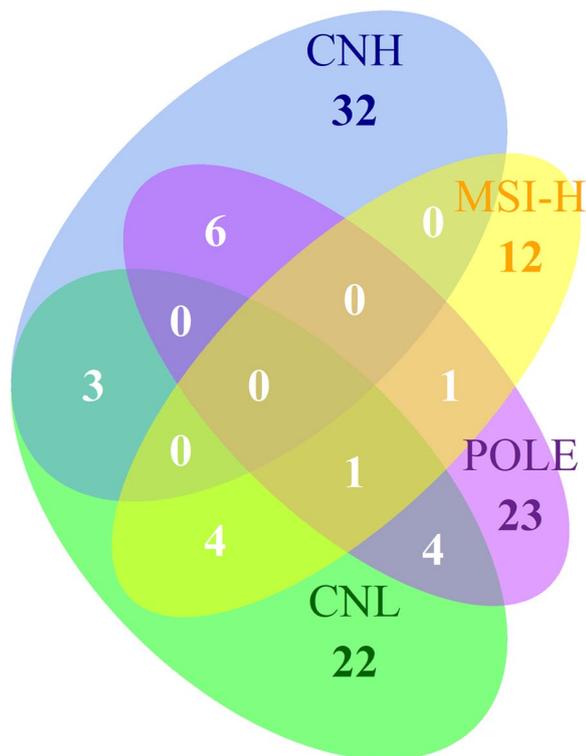


Fig. 4 DEERGs shared between molecular subtypes were comparatively few. DEERG: differentially expressed EMT-related genes

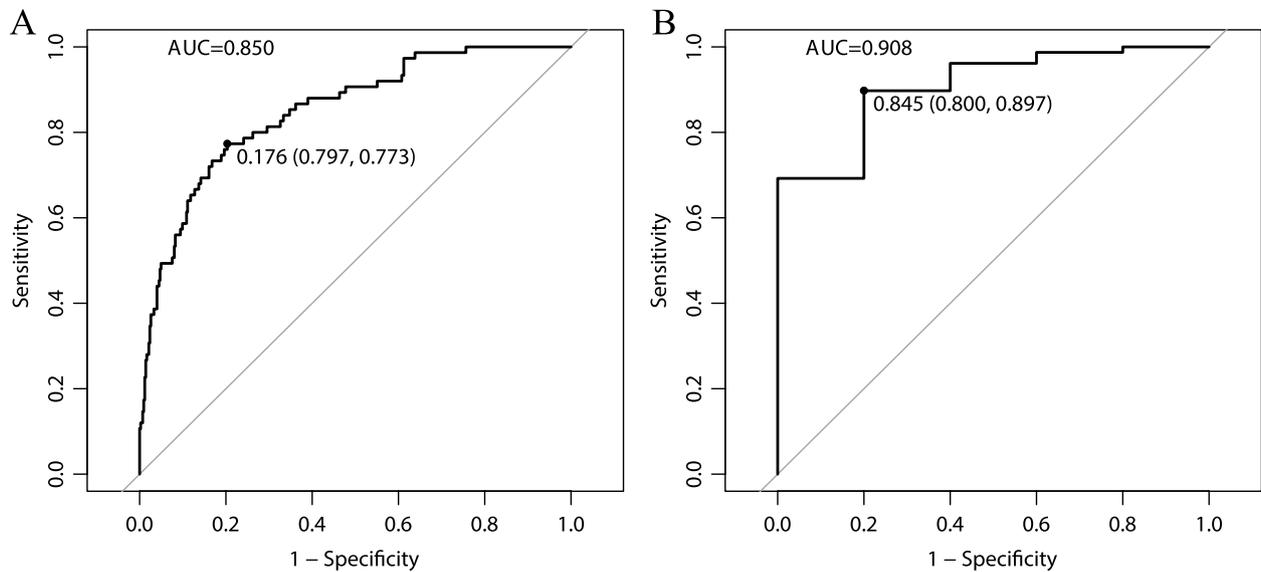


Fig. 5 A logistic regression model combining expression-based risk score and clinicopathological parameters exhibited a good discrimination in the TCGA (A) and CPTAC (B) dataset

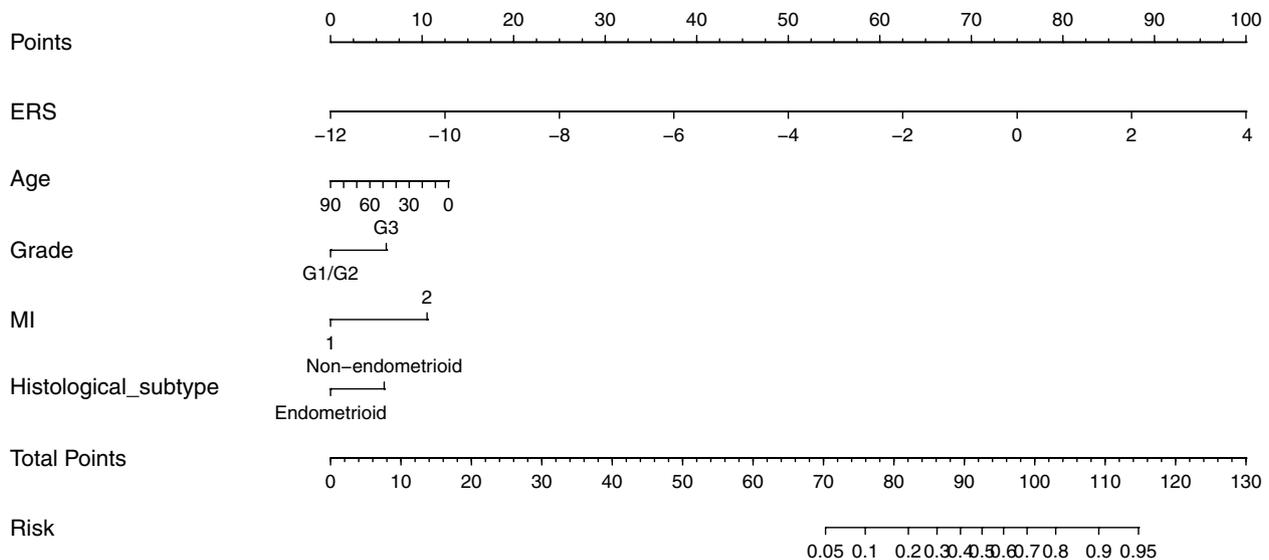


Fig. 6 ERS was the primary predictor of LNM in the nomogram

lymphatic vessels and survival in LNs [35, 36], literature concentrating on the role of EMT in LNM in EC remains few [18]. On that account, our team designed this first TCGA-based study combining transcriptomic data with conventional parameters to explore EMT-related molecular characteristics of LNM in EC.

We applied a well-validated EMT quantification method to describe a tumor sample on the EMT continuum [25]. Though unable to discriminate hybrid

E/M samples from a mixture of E&M ones, this model strongly suggested samples from LNM cases tended to have a more mesenchymal phenotype, consistent with the association between EMT and LNM in the literature, which exhibited the feasibility of further exploration of this process in EC. Intriguingly, this difference in EMT score was more significant in G3 and $MI \geq 1/2$ cases, which indicates an association between invasive phenotypes and the role of EMT in LNM. In all the molecular

subtypes, LNM cases tended to have higher EMT scores, with *POLE* showing statistical significance (Supplementary Fig. 2). This indicated a possible variation between molecular subtypes regarding the extent of dependence on EMT during LNM. There is a blank of study on EMT in different molecular subtypes, which seems promising when taking into account the abundant pathways and molecules in the EMT process [37].

Among the 1027 ERGs included in this study, over 1/4 showed a differential expression between LNM and non-LNM cases, with multiple pathways being enriched, in which the TGF- β signaling pathway has been proven a potent inducer of EMT in EC [38]. Nonetheless, no published literature discussed the role of the TGF- β pathway for LNM in EC. In the GO analysis based upon DEERGs, various pathways bound up with TGF- β were enriched in the BP group. A similar result was reached by KEGG analysis, with other EMT-related pathways significantly enriched as well, members of which were mostly upregulated (Fig. 3). Hub genes suggested by PPI analysis, e.g., *Myc*, *ESRI*, and *KRAS*, were reported to participate in EMT pathways, i.e., Wnt, TGF- β , and MAPK signaling pathways [39–44], and they also showed most interactions with other DEERGs. These hub genes might be associated with the core mechanism and containment strategy of LNM in EC in need of further exploration.

Data from this study showed different LNM rates between molecular subtypes, especially for a significantly higher rate in CNH cases, consistent with the report by Jamieson et al. [3]. DEERGs shared between subtypes were comparatively few, which exhibits possibly different landscapes of LNM between molecular subtypes in EC, as stated above. Detailed pathway analyses were omitted as a result of a small sample size in the subgroup analysis, and a prospective study on EMT in LNM and its correlation with molecular subtype in EC is being carried out in our institution (NSFC81972426).

As mentioned earlier, an accurate method of calculating the LNM probability will contribute to a precise strategy for lymphadenectomy and adjuvant therapy. Several predictive models with desirable performance have been created, predominantly based upon clinicopathological parameters [45, 46]. But before the definitive surgery, molecular features of EC can be acquired through hysteroscopy samples as well [47], providing a chance to calculate the risk of LNM more comprehensively. With the logistic regression model, we ultimately reached a gene signature composed of 7 DEERGs, the new model based upon which exhibited a satisfactory discrimination and calibration. It is also pivotal to note that in accordance with the nomogram, ERS accounted for the majority of LNM risk, which demonstrated that molecular features from the primary tumor did provide

significant information of LNM, which is independent from traditional clinicopathological parameters. This partly explained the difficulty to predict LNM in clinical practice, for traditional models (e.g., Mayo criteria) lacked paramount molecular features. In the TCGA-based logistic model, a PPV of 0.403 and an NPV of 0.952 were reached, which seems acceptable owing to the comparatively uncommon and unpredictable nature of LNM in EC. In terms of screening candidates for abandoning lymphadenectomy, a model with an NPV high enough is more crucial for medical safety [48, 49]. What is more, it is worth mentioning that SLN detection has been proven to accurately detect LNM in EC [50, 51], effectively lessen complications associated with systemic lymphadenectomy. For the time being, the number of research literature on molecular characteristics guiding SLN detection applications is still limited (predominantly molecular subtypes) [52], which evidently exhibits a promising prospect for molecular feature-guided lymphadenectomy.

The value of the 7-gene signature was further validated in a new EC database from CPTAC, where most members of the signature exhibited a positive correlation between proteins and mRNA, which conspicuously illustrates that to transform this transcriptome-based gene signature into an immunohistochemistry-based tissue microarray (TMA) from hysteroscopy biopsies is promising in clinical practice. In accordance with the latest report of Jamieson et al. on molecular subtype and LNM in EC, the molecular subtype is a potential predictor of LNM [3]. And the 7 genes proposed in this study are not covered by the commonly used Trans-PORTC and ProMisE algorithms, i.e., the 7-gene signature might provide independent predictive value preoperatively on the basis of molecular subtype. Moreover, different molecular subtypes of EC had different DEERGs in LNM (Fig. 4), which suggested the existence of different EMT pathways and variant dependence on EMT during the lymphatic invasion in EC. When EMT inhibitors are applied into clinical research to control metastasis in the future (as in anti-PD-1/PD-L1) [53], this study may contribute to precise medication in line with molecular features in EC.

There are several limitations in our research. First, owing to the heterogeneity of transcriptome from different databases, our study did not establish a universally applicable clinical prediction model. But the ERG-based gene signature did exhibit a predictive value independent of clinicopathological parameters across databases, supporting future clinical applications. Aside from that, this study is simply limited to bioinformatics analysis, not verified by specimens from our institution yet. Furthermore, the above limitations can be remedied through developing a TMA with a unified standard. Meanwhile,

there is a huge gap for laboratory research to fill on the mechanisms of EMT to induce lymphatic spread in EC.

Conclusions

To sum up, our study suggested the EMT status and expression of ERGs varied in LNM and non-LNM EC tissues, involving multiple EMT-related signaling pathways, and the distribution of DEERGs differed among molecular subtypes. Furthermore, an ERG-based gene signature including 7 DEERGs exhibited a desirable predictive value for LNM in EC, which requires further validation on the basis of clinical specimens in the future.

Abbreviations

AUC	Area under the curve
BP	Biological process
CNH	Copy-number high
CNL	Copy-number low
CPTAC	Clinical Proteomic Tumor Analysis Consortium
DAG	Directed acyclic graph
DEERG	Differentially expressed ERG
EC	Endometrial cancer
ECDF	Empirical distribution function
EEC	Endometrioid endometrial cancer
EMT	Epithelial-mesenchymal transition
EPC	Edge Percolated Component
ERG	EMT-related gene
ERS	EMT-based risk score
GO	Gene Ontology
IQR	Interquartile range
KEGG	Kyoto Encyclopedia of Genes and Genomes
LNM	Lymph node metastasis
MI	Myometrial invasion
MSI-H	Microsatellite instability-high
NPV	Negative predictive value
PPI	Protein-protein interaction
PPV	Positive predictive value
ROC	Receiver operating characteristic
SD	Standard deviation
SLND	Sentinel lymphadenectomy
TCGA	The Cancer Genome Atlas
TGF- β	Transforming growth factor-beta
TMA	Tissue microarray

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12957-023-02893-2>.

Additional file 1: Supplementary figure 1. A diagram of the mechanism of the KS method, exhibiting a tumor tissue with a more mesenchymal phenotype. The lower ECDF curve represented the higher expression of mesenchymal gene signatures, thus the EMT score of the presented sample is positive. **Supplementary figure 2.** EMT scores were distributed unevenly when stratified by grade (A), histological subtype (B), MI (C) and molecular subtypes (D). LN: lymph node; MI: myometrial invasion; CNH: Copy-number high; CNL: copy-number low; MSI-H: microsatellite instability-high. **Supplementary figure 3.** Heatmap for ERGs between tumor samples from LN positive and negative cases (normalized for exhibition). ERG: EMT-related gene; LN: lymph node. **Supplementary figure 4.** A DAG displaying the relationship between BPs from GO analysis. DAG: directed acyclic graph. BP: biological process; GO: Gene Ontology. **Supplementary figure 5.** PPI network by Cytoscape, with top 5 hub genes highlighted by colors. PPI: protein-protein interaction. **Supplementary figure 6.** Calibration plot suggesting predicted probabilities of LNM corresponded

closely to the actually observed proportions. LNM: lymph node metastasis. **Supplementary figure 7.** A significantly positive correlation between proteins and mRNA in members of the 7-gene signature based on CPTAC. X axis: mRNA, Y axis: protein. CPTAC: Clinical Proteomic Tumor Analysis Consortium.

Additional file 2: Supplementary table 1. DEERGs between LN positive and negative cases were screened out by 'limma' package.

Additional file 3: Supplementary table 2. Gene Ontology (GO) analyses were carried out by 'enrichplot' package.

Additional file 4: Supplementary table 3. Kyoto Encyclopedia of Genes and Genomes (KEGG) analyses were carried out by 'enrichplot' package.

Additional file 5: Supplementary table 4. Top 5 hub genes calculated by each method in PPI analysis.

Additional file 6: Supplementary table 5. DEERGs and their predictive value in LNM in cases with molecular subtypes.

Additional file 7: Supplementary table 6. DEERGs and their predictive value between LN positive and negative cases.

Additional file 8: Supplementary table 7. DEERGs and their predictive value in LNM when combined with clinicopathological parameters.

Additional file 9: Supplementary table 8. The 7 genes with the best performance in logistic regression were selected, with ERS based on which, both in TCGA and CPTAC database.

Acknowledgements

We thank members of the TCGA and CPTAC programs for their open-accessible databases. We also thank Dr. Ding Li from the School of Life Sciences, Peking University, for his support in bioinformatics.

Authors' contributions

LH analyzed the data and wrote the manuscript. WJ, LL, ZL, and WZ revised the manuscript. WZ designed the research and made the final revision. All authors read and approved the final manuscript.

Funding

This work was financially supported by the National Natural Science Foundation of China (81972426), Capital Health Research and Development of Special Fund (2022-2Z-4086) to Wang Zhiqi.

Availability of data and materials

All data of this research are from public databases, and algorithms and annotated data in this research can be requested from the corresponding author on the basis of academic collaboration.

Declarations

Ethics approval and consent to participate

All data are from public databases and an ethics approval is waived.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Received: 25 July 2022 Accepted: 10 January 2023

Published online: 22 February 2023

References

- Xia C, Dong X, Li H, Cao M, Sun D, He S, et al. Cancer statistics in China and United States, 2022: profiles, trends, and determinants. *Chin Med J (Engl)*. 2022;135(5):584–90.

2. Ignatov A, Lebius C, Ignatov T, Ivros S, Knueppel R, Papatthemelis T, et al. Lymph node micrometastases and outcome of endometrial cancer. *Gynecol Oncol*. 2019;154(3):475–9.
3. Jamieson A, Thompson EF, Huvila J, Leung S, Lum A, Morin C, et al. Endometrial carcinoma molecular subtype correlates with the presence of lymph node metastases. *Gynecol Oncol*. 2022;165(2):376–84.
4. García-Silva S, Benito-Martín A, Nogués L, Hernández-Barranco A, Mazariegos MS, Santos V, et al. Melanoma-derived small extracellular vesicles induce lymphangiogenesis and metastasis through an NGFR-dependent mechanism. *Nat Cancer*. 2021;2(12):1387–405.
5. Girling JE, Donoghue JF, Lederman FL, Cann LM, Achen MG, Stackler SA, et al. Vascular endothelial growth factor-D over-expressing tumor cells induce differential effects on uterine vasculature in a mouse model of endometrial cancer. *Reprod Biol Endocrinol RBE*. 2010;8:84.
6. Takahashi K, Mizukami H, Saga Y, Takei Y, Urabe M, Kume A, et al. Suppression of lymph node and lung metastases of endometrial cancer by muscle-mediated expression of soluble vascular endothelial growth factor receptor-3. *Cancer Sci*. 2013;104(8):1107–11.
7. Ma J, Li D, Kong FF, Yang D, Yang H, Ma XX. miR-302a-5p/367-3p-HMGA2 axis regulates malignant processes during endometrial cancer development. *J Exp Clin Cancer Res CR*. 2018;37(1):19.
8. Coffelt SB, Kersten K, Doornebal CW, Weiden J, Vrijland K, Hau CS, et al. IL-17-producing $\gamma\delta$ T cells and neutrophils conspire to promote breast cancer metastasis. *Nature*. 2015;522(7556):345–8.
9. Kim S, Cho H, Hong SO, Oh SJ, Lee HJ, Cho E, et al. LC3B upregulation by NANOG promotes immune resistance and stem-like property through hyperactivation of EGFR signaling in immune-refractory tumor cells. *Autophagy*. 2021;17(8):1978–97.
10. Wang C, Li Y, Jia L, Kim JK, Li J, Deng P, et al. CD276 expression enables squamous cell carcinoma stem cells to evade immune surveillance. *Cell Stem Cell*. 2021;28(9):1597–613.e7.
11. Xu K, Wang R, Xie H, Hu L, Wang C, Xu J, et al. Single-cell RNA sequencing reveals cell heterogeneity and transcriptome profile of breast cancer lymph node metastasis. *Oncogenesis*. 2021;10(10):66.
12. Zhou C, Wei W, Ma J, Yang Y, Liang L, Zhang Y, et al. Cancer-secreted exosomal miR-1468-5p promotes tumor immune escape via the immunosuppressive reprogramming of lymphatic vessels. *Mol Ther*. 2021;29(4):1512–28.
13. Chen HY, Chiang YF, Huang JS, Huang TC, Shih YH, Wang KL, et al. Isoliquiritigenin Reverses Epithelial-Mesenchymal Transition Through Modulation of the TGF- β /Smad Signaling Pathway in Endometrial Cancer. *Cancers*. 2021;13(6):1236.
14. Ciesielski P, Jóźwiak P, Forma E, Krześlak A. TET3- and OGT-Dependent Expression of Genes Involved in Epithelial-Mesenchymal Transition in Endometrial Cancer. *Int J Mol Sci*. 2021;22(24):13239.
15. Liu Y, Zhao R, Chi S, Zhang W, Xiao C, Zhou X, et al. UBE2C Is Upregulated by Estrogen and Promotes Epithelial-Mesenchymal Transition via p53 in Endometrial Cancer. *Mol Cancer Res*. 2020;18(2):204–15.
16. Makker A, Goel MM. Tumor progression, metastasis, and modulators of epithelial-mesenchymal transition in endometrioid endometrial carcinoma: an update. *Endocr Relat Cancer*. 2016;23(2):R85–r111.
17. Van den Eynde C, De Clercq K, Van Bree R, Luyten K, Annibaldi D, Amant F, et al. TRP channel expression correlates with the epithelial-mesenchymal transition and high-risk endometrial carcinoma. *Cell Mol Life Sci*. 2021;79(1):26.
18. Suh DH, Kim MA, Kim HS, Chung HH, Park NH, Song YS, et al. L1 cell adhesion molecule expression is associated with pelvic lymph node metastasis and advanced stage in diabetic patients with endometrial cancer: a matched case control study. *J Cancer Prev*. 2014;19(3):231–9.
19. Giannini A, Bogani G, Vizza E, Chiantera V, Laganà AS, Muzii L, et al. Advances on Prevention and Screening of Gynecologic Tumors: Are We Stepping Forward? *Healthcare*. 2022;10(9):1605.
20. Sun H, Chang Q, Liu YS, Jiang YT, Gong TT, Ma XX, et al. Adherence to Cancer Prevention Guidelines and Endometrial Cancer Risk: Evidence from a Systematic Review and Dose-Response Meta-analysis of Prospective Studies. *Cancer Res Treat*. 2021;53(1):223–32.
21. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov*. 2012;2(5):401–4.
22. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal*. 2013;6(269):pl1.
23. Goldman MJ, Craft B, Hastie M, Repecka K, McDade F, Kamath A, et al. Visualizing and interpreting cancer genomics data via the Xena platform. *Nat Biotechnol*. 2020;38(6):675–8.
24. Tan TZ, Miow QH, Miki Y, Noda T, Mori S, Huang RY-J, et al. Epithelial-mesenchymal transition spectrum quantification and its efficacy in deciphering survival and drug responses of cancer patients. *EMBO Mol Med*. 2014;6(10):1279–93.
25. Chakraborty P, George JT, Tripathi S, Levine H, Jolly MK. Comparative Study of Transcriptomics-Based Scoring Metrics for the Epithelial-Hybrid-Mesenchymal Spectrum. *Front Bioeng Biotechnol*. 2020;8:220.
26. Zhao M, Liu Y, Zheng C, Qu H. dbEMT 2.0: An updated database for epithelial-mesenchymal transition genes with experimentally verified information and precalculated regulation information for cancer metastasis. *J Genet Genomics*. 2019;46(12):595–7.
27. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. 2015;43(7):e47.
28. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res*. 2019;47(D1):D607–13.
29. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*. 2003;13(11):2498–504.
30. Chin CH, Chen SH, Wu HH, Ho CW, Ko MT, Lin CY. cytoHubba: identifying hub objects and sub-networks from complex interactome. *BMC Syst Biol*. 2014;8 Suppl 4(Suppl 4):S11.
31. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*. 2011;12:77.
32. Dou Y, Kawaler EA, Cui Zhou D, Gritsenko MA, Huang C, Blumenberg L, et al. Proteogenomic Characterization of Endometrial Carcinoma. *Cell*. 2020;180(4):729–48.e26.
33. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2022. URL <https://www.R-project.org/>.
34. Li L, Tang M, Nie D, Gou J, Li Z. Para-aortic lymphadenectomy did not improve overall survival among women with type I endometrial cancer. *Int J Gynaecol Obstet*. 2020;150(2):163–8.
35. Pang MF, Georgoudaki AM, Lambut L, Johansson J, Tabor V, Hagikura K, et al. TGF- β 1-induced EMT promotes targeted migration of breast cancer cells through the lymphatic system by the activation of CCR7/CCL21-mediated chemotaxis. *Oncogene*. 2016;35(6):748–60.
36. Li X, Tang J, Du H, Wang X, Wu L, Hu P, et al. Immune Characters and Plasticity of the Sentinel Lymph Node in Colorectal Cancer Patients. *J Immunol Res*. 2021;2021:5516399.
37. Babaei G, Aziz SG-G, Jaghi NZZ. EMT, cancer stem cells and autophagy; The three main axes of metastasis. *Biomed Pharmacother*. 2021;133:110909.
38. Sahoo SS, Quah MY, Nielsen S, Atkins J, Au GG, Cairns MJ, et al. Inhibition of extracellular matrix mediated TGF- β signalling suppresses endometrial cancer metastasis. *Oncotarget*. 2017;8(42):71400–17.
39. Li C, Li W, Cao S, Xu J, Qian Y, Pan X, et al. Circ_0058106 promotes proliferation, metastasis and EMT process by regulating Wnt2b/ β -catenin/c-Myc pathway through miR-185-3p in hypopharyngeal squamous cell carcinoma. *Cell Death Dis*. 2021;12(11):1063.
40. Tian M, Schiemann WP. TGF- β Stimulation of EMT Programs Elicits Non-genomic ER- α Activity and Anti-estrogen Resistance in Breast Cancer Cells. *J Cancer Metastasis Treat*. 2017;3:150–60.
41. Shao DD, Xue W, Krall EB, Bhutkar A, Piccioni F, Wang X, et al. KRAS and YAP1 converge to regulate EMT and tumor survival. *Cell*. 2014;158(1):171–84.
42. Pan LN, Ma YF, Li Z, Hu JA, Xu ZH. KRAS G12V mutation upregulates PD-L1 expression via TGF- β /EMT signaling pathway in human non-small-cell lung cancer. *Cell Biol Int*. 2021;45(4):795–803.

43. Buonato JM, Lazzara MJ. ERK1/2 blockade prevents epithelial-mesenchymal transition in lung cancer cells and promotes their sensitivity to EGFR inhibition. *Can Res.* 2014;74(1):309–19.
44. Qiao Y, Wang Z, Tan F, Chen J, Lin J, Yang J, et al. Enhancer Reprogramming within Pre-existing Topologically Associated Domains Promotes TGF- β -Induced EMT and Cancer Metastasis. *Mol Ther.* 2020;28(9):2083–95.
45. Li X, Cheng Y, Dong Y, Zhou J, Wang Z, Li X, et al. Development and validation of predictive model for lymph node metastasis in endometrial cancer: a SEER analysis. *Ann Transl Med.* 2021;9(7):538.
46. Berg HF, Ju Z, Myrvold M, Fasmer KE, Halle MK, Hoivik EA, et al. Development of prediction models for lymph node metastasis in endometrioid endometrial carcinoma. *Br J Cancer.* 2020;122(7):1014–22.
47. Talhouk A, Hoang LN, McConechy MK, Nakonechny Q, Leo J, Cheng A, et al. Molecular classification of endometrial carcinoma on diagnostic specimens is highly concordant with final hysterectomy: Earlier prognostic information to guide treatment. *Gynecol Oncol.* 2016;143(1):46–53.
48. Kang S, Thompson Z, McClung EC, Abdallah R, Lee JK, Gonzalez-Bosquet J, et al. Gene Expression Signature-Based Prediction of Lymph Node Metastasis in Patients With Endometrioid Endometrial Cancer. *Int J Gynecol Cancer.* 2018;28(2):260–6.
49. Kang S, Nam J-H, Bae D-S, Kim J-W, Kim M-H, Chen X, et al. Preoperative Assessment of Lymph Node Metastasis in Endometrial Cancer: A Korean Gynecologic Oncology Group Study. *Cancer.* 2017;123(2):263–72.
50. Persson J, Salehi S, Bollino M, Lönnerfors C, Falconer H, Geppert B. Pelvic Sentinel lymph node detection in High-Risk Endometrial Cancer (SHREC-trial)-the final step towards a paradigm shift in surgical staging. *Eur J Cancer.* 2019;116:77–85.
51. Hagen B, Valla M, Aune G, Ravlo M, Abusland AB, Araya E, et al. Indocyanine green fluorescence imaging of lymph nodes during robotic-assisted laparoscopic operation for endometrial cancer. A prospective validation study using a sentinel lymph node surgical algorithm. *Gynecol Oncol.* 2016;143(3):479–83.
52. Concin N, Matias-Guiu X, Vergote I, Cibula D, Mirza MR, Marnitz S, et al. ESGO/ESTRO/ESP guidelines for the management of patients with endometrial carcinoma. *Int J Gynecol Cancer.* 2021;31(1):12–39.
53. Musacchio L, Boccia SM, Caruso G, Santangelo G, Fischetti M, Tomao F, et al. Immune Checkpoint Inhibitors: A Promising Choice for Endometrial Cancer Patients? *J Clin Med.* 2020;9(6):1721.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

