



OPEN ACCESS

EDITED BY

Liang Cheng,
Harbin Medical University, China

REVIEWED BY

Kebo LV,
Ocean University of China, China
Taigang Liu,
Shanghai Ocean University, China

*CORRESPONDENCE

Yi Shi
Shiyi.veals@qq.com
Xiaoli Shi
shixl@geneis.cn

SPECIALTY SECTION

This article was submitted to
Cancer Genetics,
a section of the journal
Frontiers in Oncology

RECEIVED 30 June 2022

ACCEPTED 03 August 2022

PUBLISHED 19 August 2022

CITATION

Li L, Qiu W, Lin L, Liu J, Shi X and Shi Y
(2022) Predicting recurrence and
metastasis risk of endometrial
carcinoma *via* prognostic signatures
identified from multi-omics data.
Front. Oncol. 12:982452.
doi: 10.3389/fonc.2022.982452

COPYRIGHT

© 2022 Li, Qiu, Lin, Liu, Shi and Shi. This
is an open-access article distributed
under the terms of the [Creative
Commons Attribution License \(CC BY\)](#).
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.

Predicting recurrence and metastasis risk of endometrial carcinoma *via* prognostic signatures identified from multi-omics data

Ling Li¹, Wenjing Qiu², Liang Lin¹, Jinyang Liu^{2,3},
Xiaoli Shi^{2,3*} and Yi Shi^{4*}

¹Department of Gynecological Oncology Surgery, Fujian Cancer Hospital, Fujian Medical University Cancer Hospital, Fuzhou, China, ²Science System Department, Geneis Beijing Co., Ltd., Beijing, China, ³Qingdao Geneis Institute of Big Data Mining and Precision Medicine, Qingdao, China, ⁴Department of Molecular Pathology, Fujian Cancer Hospital, Fujian Medical University Cancer Hospital, Fuzhou, China

Objectives: Endometrial carcinoma (EC) is one of the three major gynecological malignancies, in which 15% - 20% patients will have recurrence and metastasis. Though there are many studies on the prognosis on this cancer, the performances of existing models evaluating the risk of its recurrence and metastasis are yet to be improved. In addition, a comprehensive multi-omics analyses on the prognostic signatures of EC are on demand. In this study, we aimed to construct a relatively stable and reliable model for predicting recurrence and metastasis of EC. This will help determine the risk level of patients and choose appropriate adjuvant therapy, thereby avoiding improper treatment, and improving the prognosis of patients.

Methods: The mRNA, microRNA (miRNA), long non-coding RNA (lncRNA), copy number variation (CNV) data and clinical information of patients with EC were downloaded from The Cancer Genome Atlas (TCGA). Differential expression analyses were performed between the recurrence or metastasis group and the non-recurrence/metastasis group. Then, we screened potential prognostic markers from the four kinds of omics data respectively and established prediction models using three classifiers.

Results: We achieved differential expressed mRNAs, lncRNAs, miRNAs and CNVs between the two groups. According to feature selection scores by the random forest algorithm, 275 CNV features, 50 lncRNA features, 150 miRNA features and 150 mRNA features were selected, respectively. And the prediction model constructed by the features of lncRNA data using random forest method showed the best performance, with an area under the curve of 0.763, and an accuracy of 0.819 under 10-fold cross-validation.

Conclusion: We developed a computational model using omics information, which is able to predicting recurrence and metastasis risk of EC accurately.

KEYWORDS

endometrial carcinoma, recurrence and metastasis, lncRNA, CNV, mRNA, miRNA, prediction model

Introduction

Endometrial carcinoma (EC) is a kind of epithelial malignant tumor occurring in the endometrium and is one of the three major gynecological malignancies (1–3). In North America and Europe, it is the fourth leading cancer following breast cancer, lung cancer, colorectal tumor in terms of incidence (4). In China, the incidence of the disease is also increasing year by year and is second only to cervical cancer (5). Obesity, hormonal and metabolic disorders are particularly closely related to the occurrence of EC (6). Its clinical treatment is mainly surgical resection, supplemented by radiotherapy and drug treatment. Although most patients are at the early stage when diagnosed and have a good prognosis, 15% - 20% of patients will have recurrence and metastasis (7–9). The presence of poor prognosis of recurrence or metastasis is the main cause leading to the death of EC patients (10, 11). Therefore, accurate prediction of the recurrence and metastasis of endometrial cancer as early as possible and performed targeted adjuvant therapy are essential to improve the survival rate of EC patients. In fact, it is difficult to identify patients with a high risk of recurrence and metastasis in the early stage. Traditionally, clinicians usually predict the risk of recurrence and metastasis by pathological type, histological grade, depth of myometrial invasion, lymphatic metastasis and extrauterine lesions, and monitor the development of the disease through patients' regular radiologic examination and laboratory examinations (12–15).

Nowadays, with the development of liquid biopsy technology and the popularization of artificial intelligence in the field of medical images, there are many new explorations and novel methods in predicting tumor recurrence and metastasis (16–22). For example, Wu et al. developed a deep convolutional neural network (CNN) model to predict the risk of recurrence

and metastasis from hematoxylin and eosin (H&E) stained sections of lung cancer (23). For estimating the risk of recurrence and metastasis in patients with HER2-positive breast cancer, Yang et al. constructed a novel multimodal fusion model integrating H&E images and clinical characteristics (20), with an area under the curve (AUC) of 0.72 in the independent testing data. Feng et al. identified that detection of somatic mutations of ctDNA could predict recurrence of EC effectively and stably (16). Ye et al. developed a deep convolution network to predict cervical cancer metastasis and recurrence risk (24). Based on the study results of The Cancer Genome Atlas (TCGA), endometrial cancer was classified into four categories according to the mutation spectrum, somatic copy number alterations (SCNAs) and microsatellite instability (MSI): DNA polymerase epsilon (POLE) ultramutated, high microsatellite instability (MSI-H), copy-number low, and copy-number high (25). TCGA molecular typing has initially shown good application prospects in predicting the prognosis of endometrial cancer patients and has been listed in the national comprehensive cancer network (NCCN), which may affect post-surgical adjuvant treatment. However, no applicable prognostic prediction models only based on genomics have been found by retrieving concerned literatures (8, 10).

In this study, all sequencing data and clinical information of patients with EC from TCGA (<http://cancergenome.nih.gov/>) were downloaded and organized to study the association between gene mutation/expression and recurrence or metastasis of EC. Specifically, we first compared the differential expressions of mRNA, long non-coding RNA (lncRNA) and microRNA (miRNA) between patients with recurrence or metastasis and patients without recurrence or metastasis using the DESeq2 package, and then analyzed differences of copy number variations (CNVs) between the two groups by rank-sum test. Furthermore, we analyzed the function of these differential genes, and discussed the molecular mechanism of recurrence and metastasis of EC. After that, characteristic variables were selected by a random forest (RF) algorithm with feature selection and were used to establish prognostic prediction models using three different classifiers. Finally, the RF model based on lncRNA showed the best performance among the twelve models.

Abbreviations: EC, endometrial carcinoma; lncRNAs, long non-coding RNAs; CNV, copy number variation; NCCN, national comprehensive cancer network; TCGA, The Cancer Genome Atlas; AUC, area under the curve; ROC, receiver operating characteristic; RF, random forest; GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes.

Materials and methods

Study participants

TCGA is a great cancer genome project which has produced genomic, epigenomic, transcriptomic and proteomic data of more than 20,000 cancer patients covering multiple cancer types. These data can help researchers have a more comprehensive understanding of cancer and improve the level of cancer screening, diagnosis and treatment. Clinical information of 548 patients with EC was downloaded from the TCGA data portal, including 204 patients without recurrence or metastasis, 43 patients with recurrence or metastasis and 301 patients without information of recurrence and metastasis. In addition, mRNA sequencing data of 543 EC patients, miRNA data of 538 EC patients, lncRNA data of 537 EC patients and CNV data of 534 EC patients were downloaded. Then, we matched the data according to the patient ID, and selected patients with complete omics data and prognostic information into the study.

Difference analysis

The expression data of mRNAs, lncRNAs and miRNAs were displayed as reads per million (RPM) and the expression levels were normalized by DESeq2 (26) package of R language for difference analysis. Then the differentially expressed mRNAs, lncRNAs and miRNAs were calculated by DESeq2 with $P_{adj} < 0.05$ and the absolute $\log_2FC > 1$ as the cutoff value, respectively. The CNVs of two groups (recurrence and metastasis group and non-recurrence/metastasis group) were analyzed by SPSS statistical software and significant differences were screened by rank-sum test with $P < 0.005$ as the threshold.

To explore the potential biological functions of these differential genes and the signal pathways they may participate in, we performed Gene Ontology (GO) (27, 28) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses by employing clusterProfiler R package (29) with $p_value < 0.05$ and $q_value > 1$ as the threshold.

Feature selection for modeling

The patients in the recurrence and metastasis group and in the group without recurrence or metastasis were divided into the training set and the test set with the ratio of 7:3, respectively. After the division, the patients were fixed in the training set or the test set, that is, in different omics analyses, the same patient was always in the training set or test set.

For each omics data of the training set, a feature selection algorithm based on RF was applied to screen important features

(30–33). Specifically, we screened the characteristic variables with scores according to Gini index. Then the features were grouped in steps of 25 and performed 10 fold cross-validation and scored to confirm the final number of features.

Model construction and comparison

Based on the selected characteristic parameters, RF, logistic regression and support vector machine classifiers were chosen for model construction to select the best model to predict the recurrence or metastasis of patients with endometrial cancer. Specifically, omics data in the training set were grid searched in each classifier to select the best super parameter, and then the final super parameter was determined through 10 fold cross-validation. Finally, we obtained 12 prediction models and compared the prediction performance mainly using the AUC of receiver operating characteristic (ROC) curves, precision and accuracy.

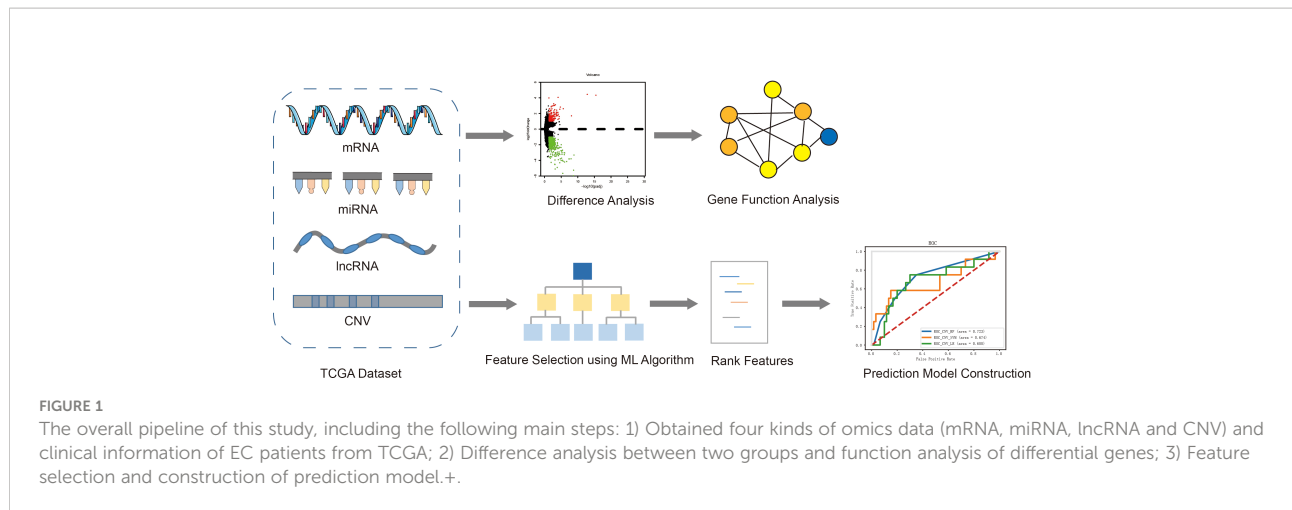
Results

A brief study design of exploring molecular mechanism and establishing prediction model

The overall process of exploring the molecular mechanism of recurrence and metastasis, and establishing risk prediction models using a machine learning algorithm was described in Figure 1. Firstly, four kinds of omics data and clinical information of EC patients were downloaded from TCGA, and then the patients were divided into two groups according to the prognosis status. Secondly, differential expression analysis was performed between the recurrent and metastatic group and non-recurrent metastatic group. Furthermore, we analyzed the function of differential genes using GO and KEGG. At the same time, characteristic variables were selected by a RF algorithm with feature selection and were used to establish prognostic prediction models using three different classifiers.

Clinicopathological features of patients with EC

After matching data according to the patient ID, 238 EC patients with both prognostic information and four kinds of omics data were obtained. Of these patients, 39 patients (16.39%) had cancer recurrence or metastasis whereas 199 patients (83.61%) had no recurrence or metastasis. Clinical and pathological information of these two groups of patients in this study was shown in Table 1. Because some information is absent, such as lymph node, progesterone receptor and estrogen



receptor status, only some factors that may be related to the prognosis of patients (7, 34) were selected for statistical analysis. As can be seen, clinical stage was significantly associated with recurrence or metastasis in this set of data, with a P -value of 0.000656. Whereas, there were no significant differences between the recurrent or metastatic group and non-recurrent or metastatic group in ages (median), body mass index (BMI) and pathological type.

Results of differential expression analysis

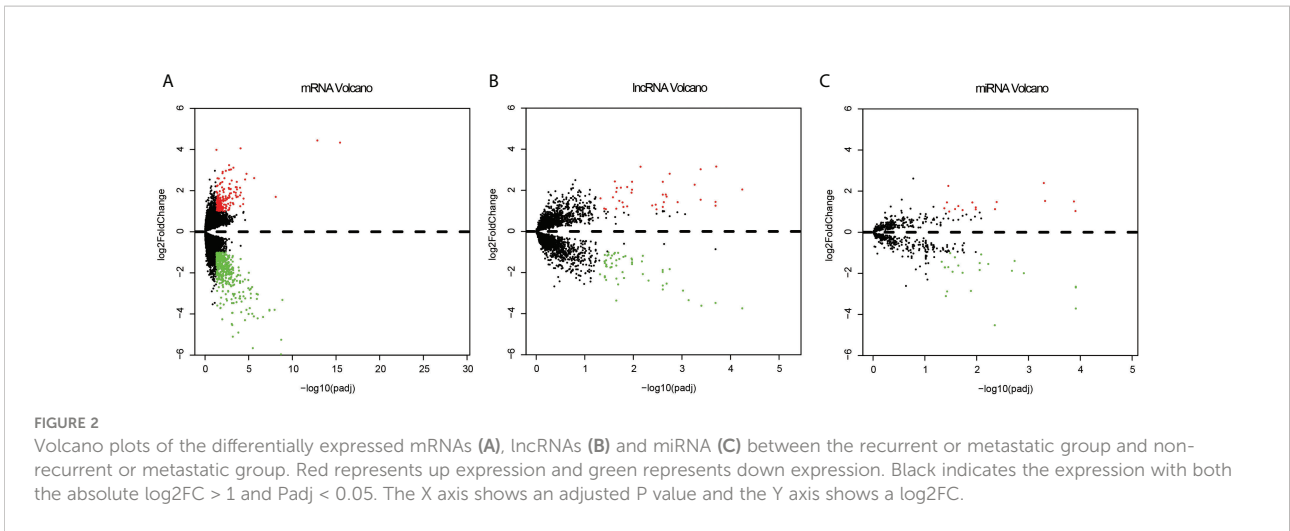
Among the expression data of 17,958 mRNAs, 592 mRNA genes were expressed significantly different between the recurrent and metastatic group and non-recurrent or metastatic group, with 169 up-regulated genes and 423 down-regulated genes in the recurrent or metastatic group compared to the non-recurrent or metastatic group (Figure 2A). And 3,352 differentially lncRNAs were achieved, in which 87 lncRNAs were significantly different, with 51 down-regulated and 36 up-regulated (Figure 2B). In addition, there were 687 differentially expressed miRNAs, in which 39 miRNAs were significantly

different, with 23 down-regulated and 16 up-regulated (Figure 2C). Heatmaps of the top 50 differentially expressed mRNA, lncRNA and miRNA were shown in Supplementary Figures 1–3. As CNVs had been reported to affect the recurrence of EC (34), it was selected separately for analysis. Finally, 939 significantly different CNVs were got after analyzing by SPSS statistical software with $P < 0.005$ as the threshold.

Then GO and KEGG enrichment analyses were performed to explore the function and involved signal pathways for further investigating the prognostic value and molecular mechanisms. For significantly differentially expressed mRNAs, the top twelve molecular functions with the highest proportion of genes were displayed in Figure 3A, and first twelve enriched signaling pathways were shown in Figure 3B. We found that the molecular functions of these differentially expressed mRNAs mainly enriched in signaling receptor activator activity, receptor ligand activity, growth factor activity, sodium ion transmembrane transporter activity, peptidase inhibitor activity, serine-type endopeptidase inhibitor activity, and so on. And the results of KEGG pathway analysis indicated that the recurrence or metastasis of EC may be correlated to the regulation of cytokine-cytokine receptor interaction, Calcium

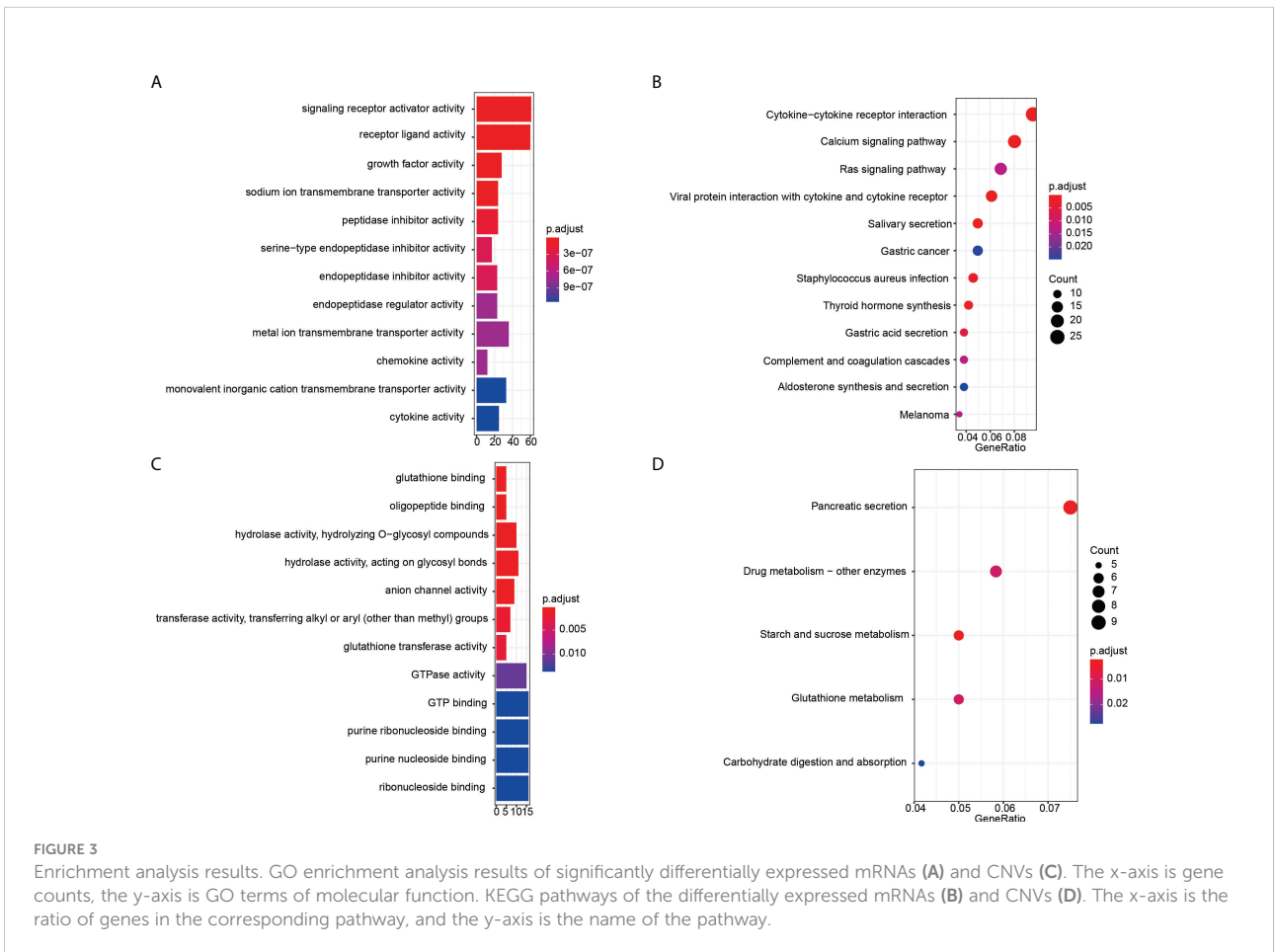
TABLE 1 Summary of clinical information of patients with EC.

Clinicopathologic variable	Category	Recurrence or metastasis	Without recurrence or metastasis	P -value
Age (years)	Median	62	61	1.000
Clinical stage	I	18	124	0.000656
	II	5	28	
	III	11	45	
	IV	5	2	
Pathological type	Endometrioid carcinoma	23	142	0.179
	Non-endometrioid carcinoma	16	57	
BMI	≤ 28	7	62	0.845
	> 28	32	137	



signaling pathway, Ras signaling pathway, viral protein interaction with cytokine and cytokine receptor. GO enrichment analysis results and KEGG pathways of the different CNVs were displayed in Figures 3C, D, respectively. From the perspective of molecular function, these mainly focus

on glutathione binding, oligopeptide binding, hydrolase activity, hydrolyzing O-glycosyl compounds, hydrolase activity, acting on glycosyl bonds, anion channel activity, transferase activity, transferring alkyl or aryl (other than methyl) groups. The genes with significant CNV differences are mainly involved in the



following five pathways: pancreatic secretion, drug metabolism - other enzymes, starch and sucrose metabolism, glutathione metabolism and carbohydrate digestion and absorption.

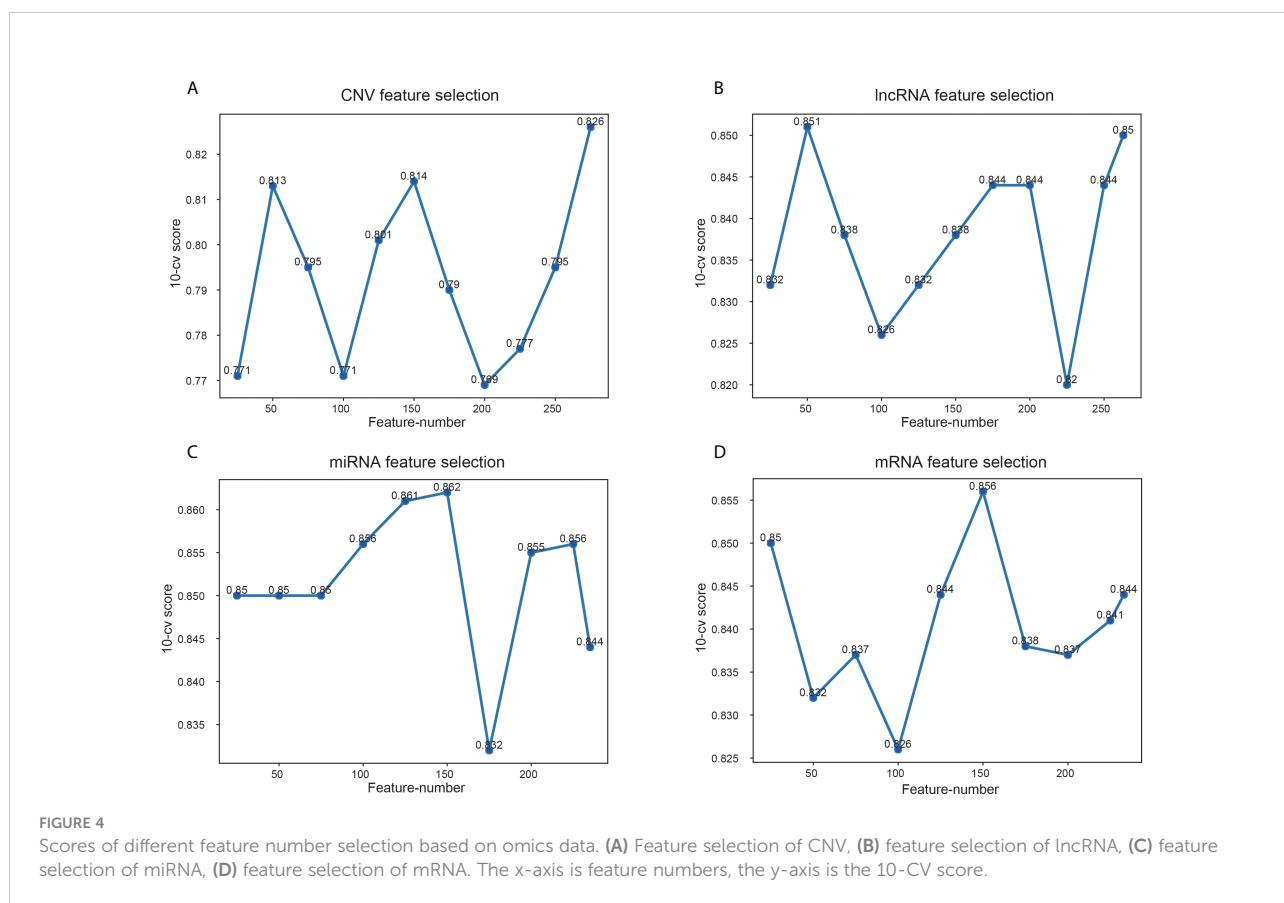
Modeling using lncRNA showed the best prediction performance

Among the data of EC downloaded from TCGA, there are 17958 mRNA expression information, 7315 lncRNA expression information, 1881 miRNA expression information and 16383 copy number variation information. Variable selection for these biological data was performed using the RF to determine variable importance measures. The scoring of different number of features screened by RF is shown in Figure 4. The features with the highest 10-CV score were selected for model construction. Specifically, 275 features were chosen from CNV data because the score of 275 features was 0.826, significantly higher than other feature combinations (Figure 4A). And 50 lncRNA features were selected as the score was 0.851, which was higher than others (Figure 4B). For the other two kinds of genomic data, 150 features of miRNA and mRNA were selected, with the highest score of 0.862 and 0.856, respectively (Figures 4C, D).

For each kind of omics data, three classifiers (RF, LR and SVM) were used to construct the prediction model. The ROC curves of three models based on lncRNA data were displayed in Figure 5A, because the model based on the characteristics of lncRNA data represented the best prediction performance. And accuracy, precision, recall and F1-score of the three models were shown in Figure 5B. The RF model constructed by the features of lncRNA data was able to predict recurrence or metastasis of EC with an AUC of 0.763, an accuracy of 0.819. The ROC curves of other models using omics variables were shown in Supplementary Figures 4. The ROC curves of models with the best prediction performance constructed by four omics data (lncRNA, mRNA, miRNA and CNV) were represented in Figure 5C, and the accuracy, precision, recall and F1-score of the prediction models were revealed in Figure 5D.

Discussion

The Oncotype Dx (21 genes) and MammaPrint (70 genes) are two products for predicting recurrence and metastasis of breast cancer which have been internationally recognized (35). However, for patients with EC, there is no effective model based on molecular variations to evaluate the risk of recurrence and



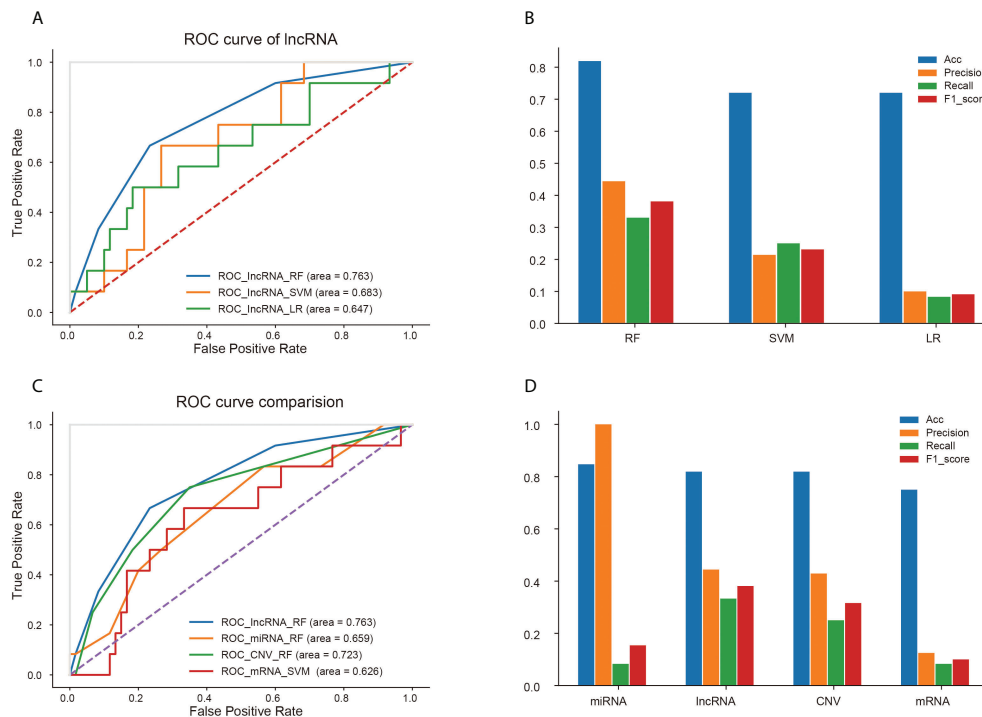


FIGURE 5

Prediction performance of different models based on four kinds of omics data. (A) ROC curves of three models based on lncRNA data. (B) Accuracy, precision, recall and F1-score of three models for lncRNA signatures. Comparison of ROC curves (C) and four properties (D) of optimal models based on four kinds of omics data.

metastasis. EC is a malignant tumor, which often occurs in perimenopausal and postmenopausal women. It usually has a good prognosis if diagnosed early and treated appropriately. So, patients will benefit greatly when a product like Oncotype DX appears, that can help clinicians assess the recurrence risk of patients and adopt adjuvant treatment strategies according to different risk stratification. Previous studies have established prediction models based on clinical characteristics (14, 15) and combined clinical characteristics with molecular data (34). AUC value of the model using clinical features only was about 0.7, whereas M. D. Miller et al. used different kinds of molecular data, up to 5 categories, it may be difficult and expensive to apply clinically.

Here, starting from the data of TCGA, we analyzed the differences of mRNA expression, miRNA expression, lncRNA expression as well as CNVs between patients with recurrence and metastasis and non-recurrence or metastasis, and further analyzed their molecular biological functions and involved signal pathways, trying to explore the molecular biological mechanism of these differences and recurrence and metastasis. Although these molecules are related to recurrence and metastasis, it does not mean that these differential molecules can accurately and reliably predict recurrence and metastasis. To build the

prediction model, it is still necessary to select the most appropriate feature combination by statistical methods (17, 34). Therefore, we used the feature selection algorithm of RF to filter features and selected different classifiers to establish models. Finally, the model using lncRNA data showed the best performance, with an AUC of 0.763, an accuracy of 0.819.

There are still several limitations in this study. Firstly, the sample size was limited and the survival time of tracking was not long enough, which may lead to inaccurate results. However, we have downloaded all samples information from TCGA, a relatively large-scale cancer genome database. Looking for more samples from other open databases or hospitals may be a solution. Secondly, the research on molecular mechanism was not deep enough (36–39). Maybe we should explore and discuss the mechanism of recurrence and metastasis of endometrial cancer in another study (40–42). Thirdly, a more advanced computational model can further improve the prediction accuracy as used elsewhere (43–46). Finally, this study lacked an independent validation set and did not develop a clinically scoring system and thresholds to discriminate risk of recurrence and metastasis. At present, we have not collected enough clinical samples for verification. We will continue to collect data to improve this study.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding authors.

Ethics statement

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

Author contributions

YS and XS designed the project. LLi, WQ, LLin, and JL collected and analyzed the data of patients with endometrial cancer. LLi and XS searched literatures and wrote the manuscript. All authors have approved the final version of the manuscript.

Conflict of interest

WQ, JL, and XS were employed by the company Geneis Beijing Co., Ltd.

References

- Amant F, Moerman P, Neven P, Timmerman D, Van Limbergen E, Vergote I. Endometrial cancer. *Lancet* (2005) 366(9484):491–505. doi: 10.1016/S0140-6736(05)67063-8
- Liu J, Zhou S, Li S, Jiang Y, Wan Y, Ma X, et al. Eleven genes associated with progression and prognosis of endometrial cancer (EC) identified by comprehensive bioinformatics analysis. *Cancer Cell Int* (2019) 19:136. doi: 10.1186/s12935-019-0859-1
- Bascuas T, Zedira H, Kropp M, Harmening N, Asrih M, Prat-Souteyrand C, et al. Human retinal pigment epithelial cells overexpressing the neuroprotective proteins PEDF and GM-CSF to treat degeneration of the neural retina. *Curr Gene Ther* (2021) 22(2):168–83. doi: 10.2174/1566523221666210707123809
- Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. *CA Cancer J Clin* (2020) 70(1):7–30. doi: 10.3322/caac.21590
- Jiang X, Tang H, Chen T. Epidemiology of gynecologic cancers in China. *J gynecologic Oncol* (2018) 29(1):e7. doi: 10.3802/jgo.2018.29.e7
- Felix AS, Yang HP, Bell DW, Sherman ME. Epidemiology of endometrial carcinoma: Etiologic importance of hormonal and metabolic influences. *Adv Exp Med Biol* (2017) 943:3–46. doi: 10.1007/978-3-319-43139-0_1
- Takahashi A, Matsuura M, Matoda M, Nomura H, Okamoto S, Kanao H, et al. Clinicopathological features of early and late recurrence of endometrial carcinoma after surgical resection. *Int J Gynecologic Cancer* (2017) 27(5):967–72. doi: 10.1097/IGC.0000000000000984
- Del Carmen MG, Boruta DM2nd, Schorge JO. Recurrent endometrial cancer. *Clin Obstet Gynecol* (2011) 54(2):266–77. doi: 10.1097/GRF.0b013e318218c6d1
- Liu H, Qiu C, Wang B, Bing P, Tian G, Zhang X, et al. Evaluating DNA methylation, gene expression, somatic mutation, and their combinations in inferring tumor tissue-of-Origin. *Front Cell Dev Biol* (2021) 9:619330. doi: 10.3389/fcell.2021.619330
- Coll-de la Rubia E, Martinez-Garcia E, Dittmar G, Gil-Moreno A, Cabrera S, Colas E. Prognostic biomarkers in endometrial cancer: A systematic review and meta-analysis. *J Clin Med* (2020) 9(6):1900. doi: 10.3390/jcm9061900
- Zhao T, Cheng L. Mutations in TREM2 change the expression levels of AD-related genes. *Ann Of Neurol* (2020) 88:S98–8. doi: 10.1016/j.jbneur.2022.01.004
- Kang SY, Cheon GJ, Lee M, Kim HS, Kim JW, Park NH, et al. Prediction of recurrence by preoperative intratumoral FDG uptake heterogeneity in endometrioid endometrial cancer. *Transl Oncol* (2017) 10(2):178–83. doi: 10.1016/j.tranon.2017.01.002
- Lee KR, Vacek PM, Belinson JL. Traditional and nontraditional histopathologic predictors of recurrence in uterine endometrioid adenocarcinoma. *Gynecol Oncol* (1994) 54(1):10–8. doi: 10.1006/gyno.1994.1158
- Senol T, Polat M, Ozkaya E, Karateke A. Tumor diameter for prediction of recurrence, disease free and overall survival in endometrial cancer cases. *Asian Pac J Cancer Prev* (2015) 16(17):7463–6. doi: 10.7314/APJCP.2015.16.17.7463

The remaining 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/fonc.2022.982452/full#supplementary-material>

SUPPLEMENTARY FIGURE 1

Heatmap of the top 50 differentially expressed mRNA.

SUPPLEMENTARY FIGURE 2

Heatmap of the top 50 differentially expressed lncRNA.

SUPPLEMENTARY FIGURE 3

Heatmap of the top 50 differentially expressed miRNA.

SUPPLEMENTARY FIGURE 4

Prediction performance of three kinds of models based on CNV data (A), miRNA data (B) and mRNA data (C).

15. Versluis MA, de Jong RA, Plat A, Bosse T, Smit VT, Mackay H, et al. Prediction model for regional or distant recurrence in endometrial cancer based on classical pathological and immunological parameters. *Br J Cancer* (2015) 113(5):786–93. doi: 10.1038/bjc.2015.268
16. Feng W, Jia N, Jiao H, Chen J, Chen Y, Zhang Y, et al. Circulating tumor DNA as a prognostic marker in high-risk endometrial cancer. *J Transl Med* (2021) 19(1):51. doi: 10.1186/s12967-021-02722-8
17. Kuhn M. Building predictive models in R using the caret package. *J Stat software* (2008) 28:1–26. doi: 10.18637/jss.v028.i05
18. Muinelo-Romay L, Casas-Arozamena C, Abal M. Liquid biopsy in endometrial cancer: New opportunities for personalized oncology. *Int J Mol Sci* (2018) 19(8):2311. doi: 10.3390/ijms19082311
19. Yang J, Hui Y, Zhang Y, Zhang M, Ji B, Tian G, et al. Application of circulating tumor DNA as a biomarker for non-small cell lung cancer. *Front Oncol* (2021) 11:725938. doi: 10.3389/fonc.2021.725938
20. Yang J, Ju J, Guo L, Ji B, Shi S, Yang Z, et al. Prediction of HER2-positive breast cancer recurrence and metastasis risk from histopathological images and clinical information via multimodal deep learning. *Comput Struct Biotechnol J* (2022) 20:333–42. doi: 10.1016/j.csbj.2021.12.028
21. He B, Dai C, Lang J, Bing P, Tian G, Wang B, et al. A machine learning framework to trace tumor tissue-of-origin of 13 types of cancer based on DNA somatic mutation. *Biochim Biophys Acta Mol Basis Dis* (2020) 1866(11):165916. doi: 10.1016/j.bbdis.2020.165916
22. Cheng L. Omics data and artificial intelligence: New challenges for gene therapy preface. *Curr Gene Ther* (2020) 20(1):1–1. doi: 10.2174/156652322001200604150041
23. Wu Z, Wang L, Li C, Cai Y, Liang Y, Mo X, et al. DeepLRHE: A deep convolutional neural network framework to evaluate the risk of lung cancer recurrence and metastasis from histopathology images. *Front Genet* (2020) 11:768. doi: 10.3389/fgene.2020.00768
24. Ye Z, Zhang Y, Liang Y, Lang J, Zhang X, Zang G, et al. Cervical cancer metastasis and recurrence risk prediction based on deep convolutional neural network. *Curr Bioinf* (2022) 17(2):164–73. doi: 10.2174/1574893616666210708143556
25. Kandath C, Schultz N, Cherniack AD, Akbani R, Liu Y, Shen H, et al. Integrated genomic characterization of endometrial carcinoma. *Nature* (2013) 497(7447):67–73. doi: 10.1038/nature12113
26. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* (2014) 15(12):550. doi: 10.1186/s13059-014-0550-8
27. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. the gene ontology consortium. *Nat Genet* (2000) 25(1):25–9. doi: 10.1038/75556
28. The Gene Ontology C. The gene ontology resource: 20 years and still GOing strong. *Nucleic Acids Res* (2019) 47(D1):D330–8. doi: 10.1093/nar/gky1055
29. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics J Integr Biol* (2012) 16(5):284–7. doi: 10.1089/omi.2011.0118
30. Deviaene M, Testelmans D, Borzee P, Buyse B, Huffel SV, Varon C. Feature selection algorithm based on random forest applied to sleep apnea detection. *Annu Int Conf IEEE Eng Med Biol Soc* (2019) 2019:2580–3. doi: 10.1109/EMBC.2019.8856582
31. Speiser JL. A random forest method with feature selection for developing medical prediction models with clustered and longitudinal data. *J BioMed Inform* (2021) 117:103763. doi: 10.1016/j.jbi.2021.103763
32. Hunt C, Montgomery S, Berkenpas JW, Sigafos N, Oakley JC, Espinosa J, et al. Recent progress of machine learning in gene therapy. *Curr Gene Ther* (2021) 22(2):132–43. doi: 10.2174/1566523221666210622164133
33. Zhao T, Hu Y, Peng J, Cheng L. DeepLGP: a novel deep learning method for prioritizing lncRNA target genes. *Bioinformatics* (2020) 36(16):4466–72. doi: 10.1093/bioinformatics/btaa428
34. Miller MD, Salinas EA, Newton AM, Sharma D, Keeney ME, Warriar A, et al. An integrated prediction model of recurrence in endometrial endometrioid cancers. *Cancer Manag Res* (2019) 11:5301–15. doi: 10.2147/CMAR.S202628
35. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: Cancer J Clin* (2021) 71(3):209–49. doi: 10.3322/caac.21660
36. Birkeland E, Wik E, Mjøs S, Hoivik EA, Trovik J, Werner HM, et al. KRAS gene amplification and overexpression but not mutation associates with aggressive and metastatic endometrial cancer. *Br J Cancer* (2012) 107(12):1997–2004. doi: 10.1038/bjc.2012.477
37. Caley DP, Pink RC, Trujillano D, Carter DR. Long noncoding RNAs, chromatin, and development. *ScientificWorldJournal* (2010) 10:90–102. doi: 10.1100/tsw.2010.7
38. Hou M, Tang X, Tian F, Shi F, Liu F, Gao G. AnnoLnc: a web server for systematically annotating novel human lncRNAs. *BMC Genomics* (2016) 17(1):931. doi: 10.1186/s12864-016-3287-9
39. Cheng L, Hu Y, Sun J, Zhou M, Jiang Q. DincRNA: a comprehensive web-based bioinformatics toolkit for exploring disease associations and ncRNA function. *Bioinformatics* (2018) 34(11):1953–6. doi: 10.1093/bioinformatics/bty002
40. Park SA, Kim LK, Kim YT, Heo TH, Kim HJ. Long non-coding RNA steroid receptor activator promotes the progression of endometrial cancer via wnt/β-catenin signaling pathway. *Int J Biol Sci* (2020) 16(1):99–115. doi: 10.7150/ijbs.35643
41. Peng L, Yuan XQ, Liu ZY, Li WL, Zhang CY, Zhang YQ, et al. High lncRNA H19 expression as prognostic indicator: data mining in female cancers and polling analysis in non-female cancers. *Oncotarget* (2017) 8(1):1655–67. doi: 10.18632/oncotarget.13768
42. Rinn JL, Chang HY. Genome regulation by long noncoding RNAs. *Annu Rev Biochem* (2012) 81:145–66. doi: 10.1146/annurev-biochem-051410-092902
43. Meng Y, Lu C, Jin M, Xu J, Zeng X, Yang J. A weighted bilinear neural collaborative filtering approach for drug repositioning. *Brief Bioinform* (2022) 23(2):bbab581. doi: 10.1093/bib/bbab581
44. Xu J, Cai L, Liao B, Zhu W, Yang J. CMF-impute: an accurate imputation tool for single-cell RNA-seq data. *Bioinformatics* (2020) 36(10):3139–47. doi: 10.1093/bioinformatics/btaa109
45. Liu C, Wei D, Xiang J, Ren F, Huang L, Lang J, et al. An improved anticancer drug-response prediction based on an ensemble method integrating matrix completion and ridge regression. *Mol Ther Nucleic Acids* (2020) 21:676–86. doi: 10.1016/j.omtn.2020.07.003
46. Tang X, Cai L, Meng Y, Xu J, Lu C, Yang J. Indicator regularized non-negative matrix factorization method-based drug repurposing for COVID-19. *Front Immunol* (2020) 11:603615. doi: 10.3389/fimmu.2020.603615