



OPEN Exploration of shared pathogenic factors and causative genes in early-stage endometrial cancer and osteoarthritis

Yiyun Bai¹, Sang Luo², Ruzhen Shuai³, Xiaomei Zhang⁴, Liwei Yuan⁵ & Dan Liu^{5,6,7}✉

Osteoarthritis (OA) has been implicated in the development and progression of early-stage endometrial cancer (EC), suggesting shared pathogenic factors between the two diseases. This study aimed to investigate the causal relationship between OA and EC and to identify causative genes common to both early-stage EC and OA. A Two-sample Mendelian randomization (MR) analysis was first performed to assess the causal relationship between OA and EC. Differentially expressed genes associated with early-stage EC and OA were identified using the limma package. Overlapping genes were extracted to determine common causative genes, followed by enrichment analysis. The causal relationship between these genes and EC was verified through Mendelian randomization (MR) of drug targets. Genes with diagnostic value were identified using multiple machine learning algorithms to construct EC prediction models and evaluate their performance. Additionally, the study examined the correlation between diagnostic-value genes and immune cell infiltration. IVW analysis indicated that OA was a high-risk factor for the development of EC ($P < 0.05$). Seven common causative genes (CDKN2A, DDA1, LRRC42, POLB, ADCYAP1R1, DNMT3A, and GLRX5) were identified for OA and EC, showing significant enrichment in related pathways such as heterochromatin. MR analysis of drug targets revealed that CDKN2A, DDA1, LRRC42, and POLB had diagnostic value for EC. The EC prediction model based on these four genes demonstrated high performance (AUC = 0.974 for the training set; AUC = 0.966 for the validation set), and these genes were significantly associated with immune cell infiltration ($P < 0.05$). CDKN2A, DDA1, LRRC42, and POLB may be common causative genes for OA and early-stage EC, potentially serving as targets for drug intervention.

Keywords Mendelian randomization, Osteoarthritis, Endometrial cancer, Enrichment analysis, Nomogram

Abbreviations

EC	Endometrial cancer
OA	Osteoarthritis
MR	Mendelian randomization
log2FC	Log2 fold change
GEO	Gene Expression Omnibus
GWAS	Genome-wide association studies
IVW	Inverse-variance weighted
eQTLs	expression quantitative trait loci

Endometrial cancer (EC) is the sixth most common cancer among women, particularly prevalent in high-income countries. Its incidence and mortality rates are rising globally¹. In 2020, there were 417,367 new confirmed

¹The First Clinical Medical College of Ningxia Medical University, Yinchuan 750004, Ningxia, China. ²Department of Beijing National Biochip Research CenterSub-Center in Ningxia, Institute of Medical Sciences, General Hospital of Ningxia Medical University, Yinchuan 750004, Ningxia, China. ³Department of Obstetrics and Gynecology, Gansu Provincial Hospital, Lanzhou 730000, China. ⁴Department of Obstetrics, General Hospital of Ningxia Medical University, Yinchuan 750004, Ningxia, China. ⁵Department of Gynecology, General Hospital of Ningxia Medical University, Yinchuan 750004, Ningxia, China. ⁶Institute of Medical Sciences, General Hospital of Ningxia Medical University, Yinchuan 750004, Ningxia, China. ⁷Key Laboratory of Ministry of Education for Fertility Preservation and Maintenance, Ningxia Medical University, Yinchuan 750004, Ningxia, China. ✉email: nxld@nyfy.com.cn

cases and 97,370 new deaths worldwide, with numbers expected to increase over the next decade. The primary symptom of EC is abnormal vaginal bleeding; however, no distinctive early changes have been identified, resulting in many patients being diagnosed at an advanced stage and missing the optimal treatment window^{1,2}. The diagnosis of EC primarily involves transvaginal ultrasound and endometrial biopsy, which are invasive^{3,4}. Thus, understanding the disease's etiology and pathogenesis is essential for improving diagnosis and treatment.

Osteoarthritis (OA) is the most common degenerative joint disease, affecting an estimated 350 million people worldwide. Its prevalence increases sharply with age, with the primary clinical manifestations being joint pain and limited mobility. Despite being diseases of different systems, EC and OA share common causative factors, such as advanced age, obesity, inflammation, and estrogens^{5–8}. For instance, adipose tissue can aromatize adrenal androgens to estrogens, which stimulate endometrial proliferation by inducing mitotic activity in endometrial cells, thereby increasing the risk of EC^{1,5}. Adipose tissue also contributes to OA pathogenesis by secreting cytokines that may influence metabolic processes in bone and joints⁹. Obesity mediates the development of EC through elevated fasting insulin levels and is also a major cause of OA. Overweight can overload the joints, leading to the destruction of articular cartilage, while obesity promotes fat deposition and insulin resistance, further contributing to OA development^{6,10}. Inflammation is also implicated in the development of both EC and OA. Interferon-induced monocyte cytokines have been positively associated with EC development⁸. Additionally, imbalances in macrophage polarization, as well as pro-inflammatory and anti-inflammatory mediators produced by macrophages, are closely associated with OA^{10,11}. A pro-inflammatory environment, characterized by elevated C-reactive protein, interleukin-6, and tumor necrosis factor- α , along with a relative lack of protective immune cell types in the endometrium, may contribute to EC development¹. Notably, one cohort study found that OA was one of the most common comorbidities of EC, with 35% of EC patients also having OA¹².

In summary, common pathogenic factors may exist between EC and OA. However, research on the shared pathogenic factors of EC and OA is scarce, with very limited evidence available. The causes of EC and OA remain unclear, and their early developmental stages exhibit considerable variability among individuals, lacking uniform typical clinical manifestations. This variability complicates the early diagnosis of both diseases. Early diagnosis is crucial for reducing the mortality rate of EC and is particularly beneficial for alleviating pain and improving the quality of life in OA patients. Thus, there is an urgent need to investigate the common pathogenic factors and potential pharmacological intervention targets for EC and OA.

MR is an approach that uses genetic variation as an instrumental variable to assess the causal relationship between environmental exposures and disease. Its central idea is Mendel's law of independent assortment, which is not susceptible to a number of confounding factors. This approach overcomes the limitations of traditional epidemiological studies and yields more reliable causal relationships¹³. Recently, MR analysis has become widely used for drug target development and drug repurposing¹⁴. The advancement of Genome-Wide Association Studies (GWAS) and molecular mechanism identification has provided a foundation for MR-based strategies, facilitating the identification of potential therapeutic targets across various diseases^{14–16}. Additionally, MR of drug targets has proven useful for modeling the pharmacological effects of drug targets in clinical trials and predicting the clinical benefits and adverse effects of treatments^{17,18}. Therefore, this study aims to investigate the causal association between EC and OA using the GWAS dataset, combined with MR analysis and transcriptome analysis, to explore potential pharmacological intervention targets. This research will offer new insights into EC and OA, provide a new direction for identifying potential drug intervention targets.

Materials and methods

Research design

This study examined the causal relationship between EC and OA using a two-sample MR method. Differential genes between EC and OA were identified, and their enrichment pathways were explored. Causal associations with EC were investigated using drug boot point MR based on the eQTLs of genes common to both conditions, with verification of consistency between the direction of causal association and the expected direction. Additionally, various algorithms were employed to identify key genes for constructing a nomogram, and the diagnostic and predictive performance for EC patients was evaluated. The study also explored the correlation between diagnostic genes and immune cell infiltration. The MR analysis method adhered to the three assumptions of MR research¹⁹ (Fig. 1) and the STRIOBE-MR principles²⁰ (Supplementary Table 1).

Data source

Data on EC and OA were obtained from the IEU Open GWAS database (<https://gwas.mrcieu.ac.uk/>), accessed on June 11, 2024 (Table 1). Bioinformatic analysis included data from The Cancer Genome Atlas (TCGA) database (EC data) and the GEO database (GSE12021-OA) (Table 2). The original studies obtained informed consent from participants, so this part of the study did not require ethics committee approval.

Screening of instrumental variables

To minimize bias from weak instrumental variables, a P -value threshold of $< 1 \times 10^{-5}$ was used as the screening criterion for strong correlation. Instrumental variables with F -statistics greater than 10 were preferentially selected. Additionally, if the intercept term of the MR-Egger regression model was not zero ($P > 0.05$), this indicated a lack of gene validity.

Instrumental variable screening for EC and OA

Setting the chain imbalance factor $r^2 = 0.001$ with a chain imbalance region width of 10,000 kb ensured that individual SNPs were independent and excluded the effect of gene pleiotropy. SNPs associated with confounders

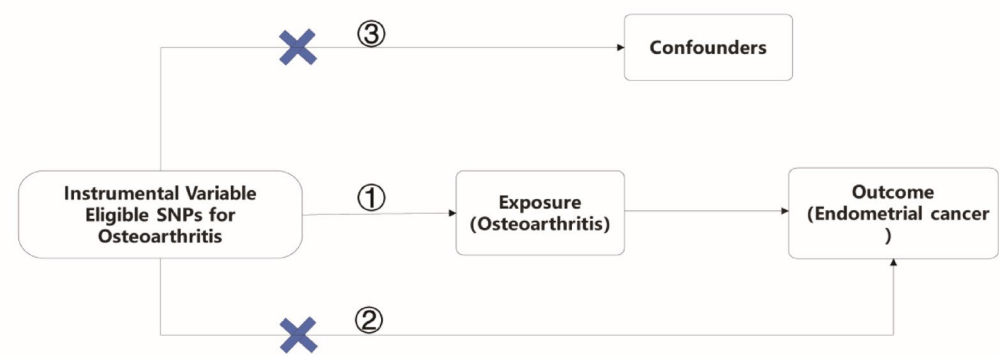


Fig. 1. Schematic diagram of MR associated with EC. Three major assumptions: ① The assumption of association: The instrumental variable is closely related to the exposure factor. ② The assumption of association: The instrumental variable is closely related to the exposure factor. ③ The assumption of independence: The instrumental variable is not correlated with confounders.

Datasource	Phenotype	Sample size	Cases	Population	Adjustment
ebi-a-GCST007092	OA of the hip or knee	417,596	39,427	European	-
ebi-a-GCST006464	EC	121,885	12,906	European	-

Table 1. Summary information on the GWAS database in the MR study.

TCGA-UCEC expression data	Excluded sample size	Remaining sample size
UCEC sample	-	554
Sample relapse	1	553
UCECStageI tissue sample	-	371
UCECStageII + III tissue sample	-	182
Normal tissue samples	-	35
Total cases	-	588
GSE12021-OA expression data	Excluded sample size	Remaining sample size
OA Sample	-	10
Control Sample	-	9
Total cases	-	19

Table 2. Inclusion and exclusion criteria.

and outcomes were removed using LDlink (<https://ldlink.nih.gov/?tab=ldtrait>).Relevant SNPs were extracted from the GWAS pooled data for EC, with a minimum $r^2 > 0.8$.

Screening of tool variables for differential genes and EC

For tool variables, the interlocking imbalance coefficient was set at $r^2 = 0.3$ with an interlocking imbalance region width of 300 kb and MAF > 0.01 to ensure SNP independence. SNPs associated with confounders and outcomes were removed using LDlink. Instrumental variables were located within ± 300 kb of the cis-acting region of the drug target gene. Relevant instrumental variables were extracted from the eQTL data of drug target genes. The SNPs were further extracted from the GWAS summary data for the outcome variable EC, excluding SNPs with palindromic structures and MAF > 0.42.SNPs directly associated with outcome variables were excluded ($P < 1 \times 10^{-5}$), and abnormal SNPs were rejected using MR-PRESSO.

MR analysis method

This study utilized five main regression models to assess the results: MR-Egger regression, random-effects inverse-variance weighted (IVW), weighted median estimator (WME), weighted and simple models. The IVW method was used as the primary analytical method to assess causality, while the MR-Egger method served as a complementary approach, particularly in the presence of horizontal pleiotropy. The presence of heterogeneity among SNPs was evaluated using Cochran's Q test and the I^2 (I-squared) statistic. Heterogeneity was indicated by a P-value ≤ 0.05 for Cochran's Q test, and $I^2 > 50\%$ suggested some heterogeneity in IVW results. The formula for $I^2 = \frac{Q - df}{Q} \times 100\%$.A non-zero intercept term ($P > 0.05$) in the MR-Egger regression model indicated that

the SNPs were not pleiotropic. The “leave-one-out” method was employed for sensitivity analyses to examine the impact of each SNP on the results and assess result robustness. All MR analyses were conducted using the Two Sample MR package in R 4.1.0 software, with a significance level of $\alpha = 0.05$.

Differential gene acquisition and functional enrichment analysis

This study used the limma (Linear Models for Microarray Data) package²¹ to perform differential expression analysis on several sets of transcriptome expression data: ① Phase I EC tissue (371 cases) versus normal control tissue (35 cases); ② Phase I EC tissues (371 cases) versus Phase II and III EC tissues (182 cases); ③ OA (10 cases) versus normal control (9 cases) tissues from the GEO dataset (GSE12021). Differential genes identified in these comparisons were intersected to find common genes. To further elucidate the potential functions of these differential genes, Differential Gene Pathway Enrichment Analysis was performed. GO (BP/CC/MF) and KEGG enrichment analyses were conducted using the R package ‘clusterProfiler’ with a q-value filter < 0.05 ^{22–24}. KEGG analysis provided insights into the high-level functions and utilities of biological systems related to differential genes, while GO analysis annotated genes based on their functions, particularly in MF, BP, and CC. The top 10 significantly enriched pathways were visualized for BP, CC, and MF enrichment analyses, and the top 30 significantly enriched pathways were visualized for KEGG enrichment analyses.

Steiger orientation test of differential genes eQTLs and EC

Common differential genes between early EC and OA were identified. GWAS summary data for eQTLs and EC (ebi-a-GCST006464) corresponding to these differential genes were obtained from the IEU Open GWAS projects database to validate the MR causal association of drug targets. Directional consistency of genotypes for causality between intermediate variables and final outcomes was assessed using the Steiger Direction Test. This method calculates the explained variance of eQTLs for the differential genes and the explained variance of EC, then tests whether the variance of EC is less than that of the differential genes. In MR Steiger results, if the variance of EC is less than the variance of the differential gene, the result is judged as ‘TRUE’, indicating that causality is in the expected direction. Conversely, a ‘FALSE’ result suggests that causality is in the opposite direction of the expected direction.

Construction and evaluation of EC prediction models

To identify candidate biomarkers and build a prediction model for EC, key genes were screened using Random Forest and LASSO regression algorithms.

Random Forest (RF) is a popular ensemble learning method that constructs predictive models with a high degree of accuracy. By building and integrating multiple decision trees, RF estimates the importance of variables, enhancing model diversity and improving the generalization of the forest. This process accurately assesses the importance of individual features. The LASSO (Least Absolute Shrinkage and Selection Operator) model is a penalized linear method used for regression analysis. It performs feature selection and regularization to prevent overfitting and improve model interpretation by adding an L1 penalty term to the loss function of ordinary least squares regression. This compresses the regression coefficients, making some coefficients zero and enabling feature selection. These techniques help reduce model complexity and enhance interpretability.

Based on the genes with diagnostic value, a nomogram was constructed using the R package ‘rms’. The area under the receiver operating characteristic (ROC) curve was plotted to evaluate the diagnostic effectiveness of these genes for EC. Finally, calibration curves and decision curve analysis (DCA) were performed to assess the efficiency of the nomogram prediction model for endometrial cancer.

Multiple hypothesis testing correction

Correction of *P*-values for MR statistics by calculating false discovery rate (FDR). It can be used as an indicator to test the error rate to flexibly adjust the value. The formula is $FDR = \frac{P_{value} \times Rank_{max}}{Rank}$.

Correlation analysis of diagnostic genes and immune cell infiltration

The diagnostic gene expression matrix was uploaded to the CIBERSORTx data library (<https://cibersortx.stanford.edu/>) to calculate immune cell infiltration for each sample. Correlation analysis between key genes and immune cell infiltration was performed using Spearman rank correlation coefficients.

Results

OA associated with increased risk of EC

A total of 92 SNPs associated with EC were identified from eQTLs data related to OA, all of which had an *F*-statistic > 10 , making them suitable as instrumental variables for assessing the causal relationship between OA and EC (Supplementary Table 2). IVW analysis indicated that OA was a significant risk factor for the development of EC (OR = 1.104, 95% CI: 1.008–1.209, $P = 0.032$). The I^2 statistic was 18%, and Cochran’s *Q* was 110.610 ($P = 0.079$), suggesting no significant heterogeneity among the SNPs used as instrumental variables in the MR analyses. MR-Egger results ($P = 0.861$) indicated that no significant pleiotropy was present among the SNPs as instrumental variables (Table 3). EC scatter plots (Fig. 2A) and funnel plots (Fig. 2C) demonstrated that the distribution of all included SNPs was generally symmetrical, suggesting that causal associations were less likely to be affected by potential bias. The leave-one-out test (Fig. 2B) showed that excluding each SNP in turn did not significantly alter the results, with no SNPs having a substantial impact on the causal association estimates, indicating that the MR results of this study are robust.

Exposure	Outcome	Nsnp	MR	P	I2(%)	Heterogeneity	P	Egger intercept	Horizontal pleiotropy	P	MR-PRESSO
			OR(95%CI)			Cochran's Q					P
OA	EC	92	1.104 (1.008–1.209)	0.032	18	110.61	0.079				0.082
OA	EC	92	1.136 (0.815–1.585)	0.454	19	110.572	0.07	-0.001	0.008	0.861	
OA	EC	92	1.091 (0.966–1.231)	0.16							
OA	EC	92	1.019 (0.740–1.402)	0.91							
OA	EC	92	1.038 (0.774–1.394)	0.803							

Table 3. Results of MR regression causal association between OA and EC.

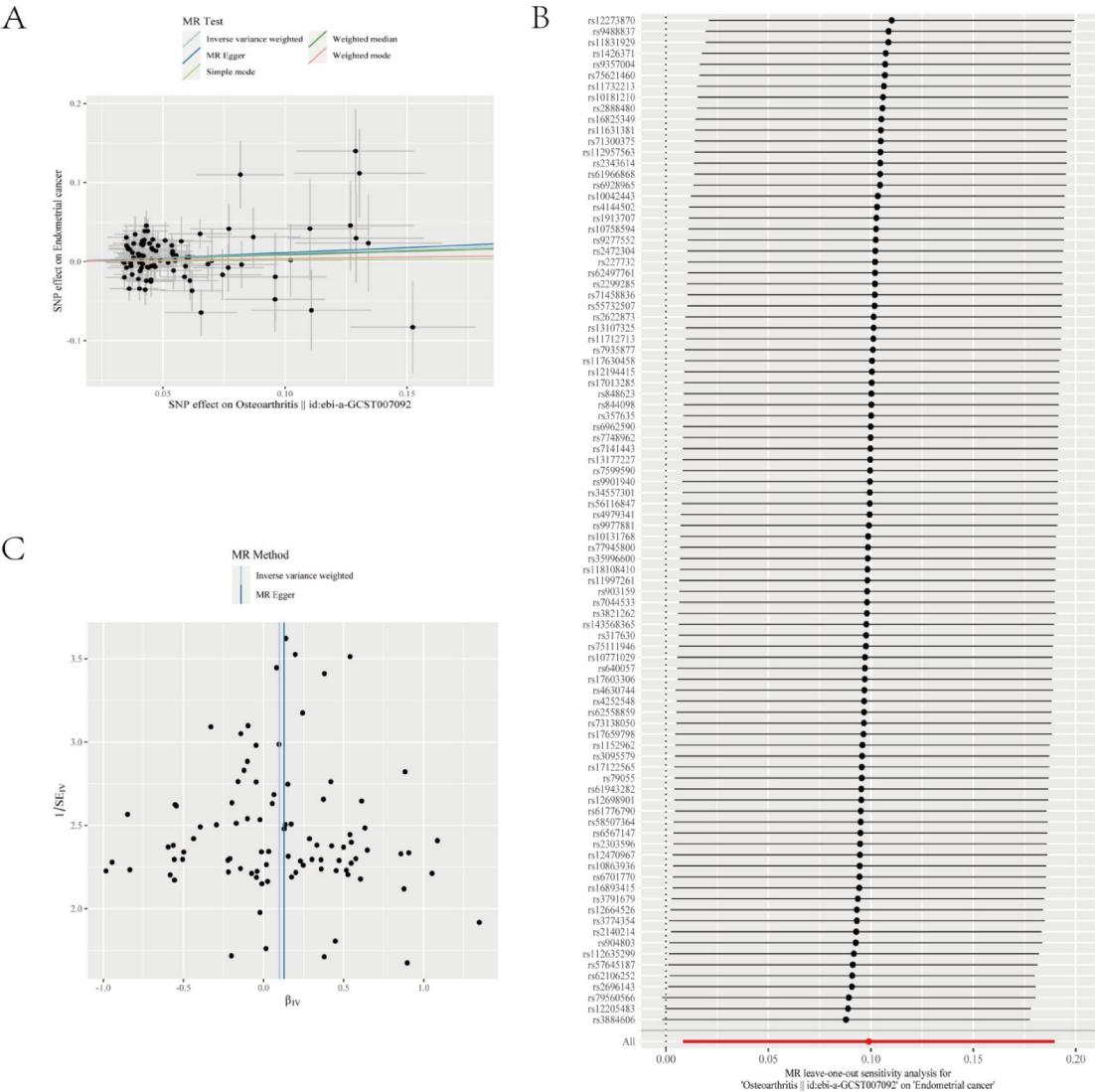


Fig. 2. Two-sample MR analysis of OA and EC. (A) Scatter plot (B) 'Leave-one-out' sensitivity analysis (C) Funnel plot.

Differential gene acquisition and functional enrichment analysis of EC and OA

Transcriptome expression data from the TCGA-UCEC dataset were analyzed for expression differences using the limma package. Initially, phase I EC was compared with normal tissue, using thresholds of $|\log_2FC| > 0.5$ and $P < 0.05$, resulting in 9,399 differential genes (Fig. 3A). Next, phase I EC tissue was compared with phase II and III EC tissue, identifying 2,607 differential genes (Fig. 3B). Intersection analysis of genes with expression differences from both comparisons, considering opposite directional changes, yielded 661 intersecting genes (Fig. 3D). Transcriptome expression data from the GSE12021 dataset were also analyzed using the limma package, with adjusted thresholds of $|\log_2FC| > 0.5$ and $P < 0.05$, resulting in 627 differentially expressed genes (Fig. 3C). Among these, seven intersecting genes with consistent expression differences across both datasets were identified (Fig. 3E).

GO enrichment analysis revealed that differential genes were significantly enriched in pathways related to heterochromatin, response to ethanol, response to alcohol, and aging. Response to ethanol, response to alcohol, and aging were associated with biological processes (BP), while heterochromatin was linked to cellular components (CC) (Fig. 3F). KEGG enrichment analysis did not identify any significant pathways.

causal association analysis of eQTLs corresponding to differential genes with EC, and Multiple hypothesis testing correction

After applying the screening criteria, seven differential genes were extracted from the IEU Open GWAS project website for six differential eQTLs (Table 4). A total of 130 cis eQTLs for these differential genes were identified from the differential eQTL data (Supplementary Table 3). IVW results indicated that four genes were associated with an increased risk of developing EC. CDKN2A (OR = 1.546, 95%CI: 1.298–1.842, $P < 0.001$) DDA1 (OR = 1.111, 95%CI: 1.021–1.208, $P = 0.014$) LRRC42 (OR = 1.112, 95%CI: 1.026–1.206, $P = 0.01$) and POLB (OR = 1.072, 95%CI: 1.004–1.145, $P = 0.038$) is a risk factor for EC (Fig. 4). There was no significant heterogeneity or pleiotropy among the instrumental variables for these three genes (Figs. S1–S3). Although the gene POLB appeared to be positively associated with EC, there may be pleiotropy among its instrumental variables, necessitating further assessment of its causality (Fig. S4). The results of the Steiger orientation test showed that the orientation of all four differential genes was ‘TRUE’, indicating that the causal relationship between these

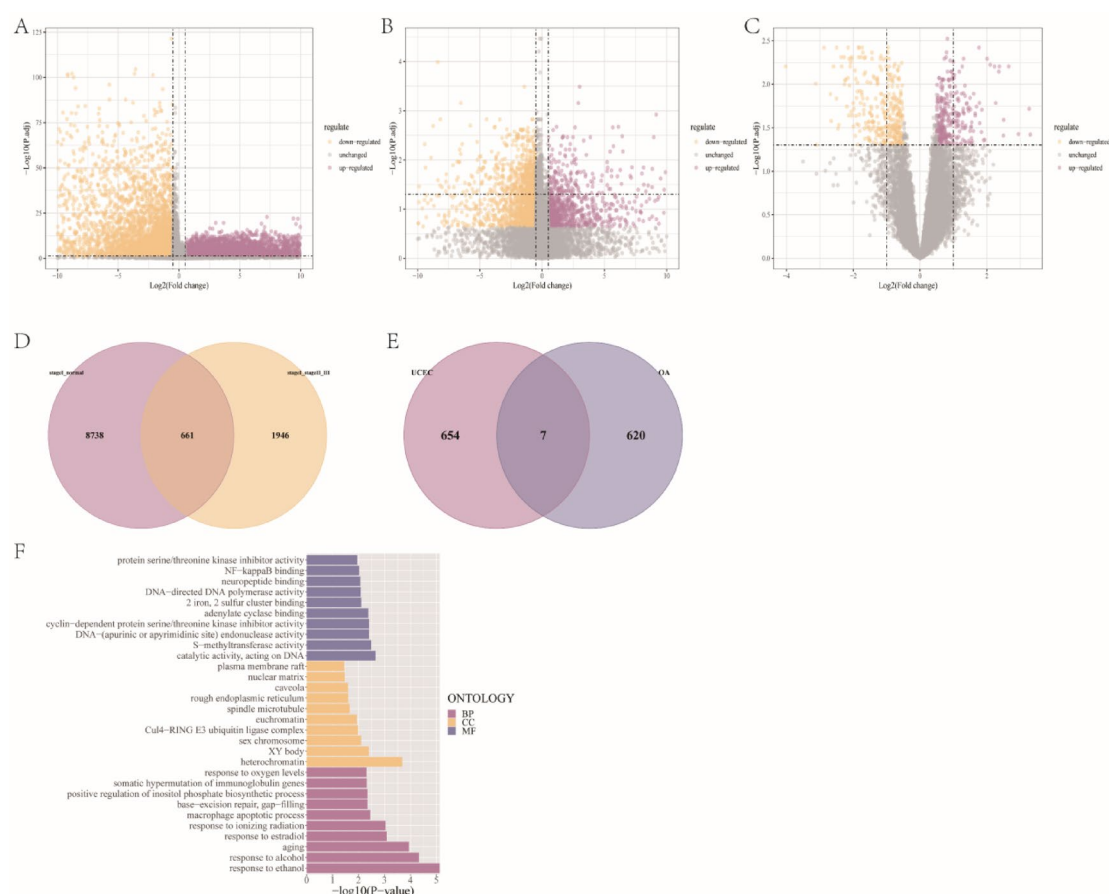


Fig. 3. Differential genes for EC and OA. (A) Differentially expressed genes in Stage I EC tissues versus normal endometrial tissues. (B) Differentially expressed genes in Stage I EC tissues versus Stage II and III EC tissues. (C) Differentially expressed genes in OA tissues versus normal tissues. (D) Genes with differential expression in opposite directions compared to Fig. 5A and Fig. 5B. (E) Genes with the same direction of differential expression as in Figure C and Figure D. (F) GO enrichment analysis.

Datasource	Phenotype	Sample size	Cases	Population	Adjustment
IEUOpenGWASproject(eqlt-a-ENSG00000070501)	POLB	31,684	-	European	MalesandFemales
IEUOpenGWASproject(eqlt-a-ENSG00000116212)	LRRC42	31,684	-	European	MalesandFemales
IEUOpenGWASproject(eqlt-a-ENSG00000119772)	DNMT3A	31,684	-	European	MalesandFemales
IEUOpenGWASproject(eqlt-a-ENSG00000147889)	CDKN2A	31,684	-	European	MalesandFemales
IEUOpenGWASproject(eqlt-a-ENSG00000182512)	GLRX5	31,684	-	European	MalesandFemales
IEUOpenGWASproject(eqlt-a-ENSG00000130311)	DDA1	14,263	-	European	MalesandFemales
IEUOpenGWASproject(ebi-a-GCST006464)	EC	121,885	12,906	European	-

Table 4. Summary information on eQTLs and GWAS databases in the MR study.

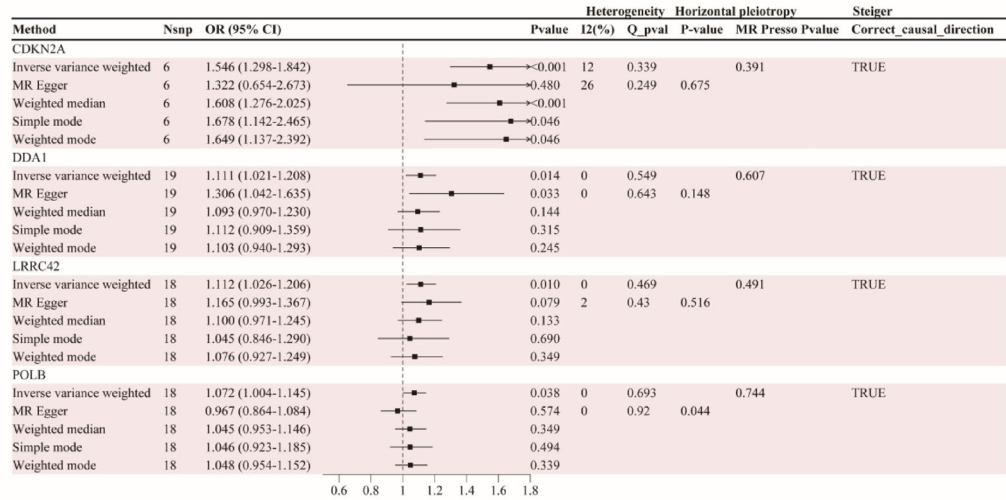


Fig. 4. Forest plot of the results of causal association analysis between differential gene eQTLs and EC.

genes and the outcome was in the expected direction (Table 5). These four genes may be common causative factors for both OA and EC, participating in and mediating the development of EC, although their specific mechanisms require further investigation. In addition, we corrected the statistics of MR for FDR, and all *P* values after this correction were less than 0.05 (Table S4).

Construction and evaluation of EC prediction models

Based on differential analysis and MR results, four genes with consistent orientation and positive analysis results were identified. These genes were screened using both the LASSO (Fig. 5A) and RF (Fig. 5B) models, resulting in the selection of four genes with diagnostic value through the intersection of the two methods (Fig. 5C). To enhance diagnostic and predictive performance for EC, a nomogram was constructed using logistic regression analysis based on these four genes (Fig. 6A). The ROC curves indicated that the model's AUC values for both the training (Fig. 6B) and validation (Fig. 6C) sets were greater than 0.96, suggesting strong diagnostic efficacy. The calibration curves demonstrated that the predictive probability of the nomogram closely matched the ideal model (Fig. 6D, E). Additionally, DCA curves indicated that decision-making based on the nomogram may improve EC diagnosis (Fig. 6F, G).

Correlation analysis of genes of diagnostic value with immune cell infiltration

Analysis of immune cell infiltration in EC revealed that the expression of the key gene CDKN2A was significantly positively correlated with the degree of NK cell activation (*P* < 0.001) and negatively correlated with the degree of activated CD4 memory T cells (*P* < 0.001). The key gene LRRC42 was significantly positively correlated with Macrophages M1, Macrophages M2, and the degree of activated Dendritic cells (*P* < 0.001). DDA1 gene expression was negatively correlated with the degree of resting CD4 memory T cells (*P* < 0.001), but positively correlated with the degree of activated NK cells (*P* < 0.01). POLB expression showed a significant positive correlation with the degree of follicular helper T cells and memory B cells (*P* < 0.01) (Fig. 7).

Discussion

In this study, we explored the potential causal relationship between OA and EC and investigated potential drug target genes for EC. MR analysis revealed small effect size (OR = 1.104) but it's still important in epidemiological

Exposure	Outcome	Nsnp	MR	P	I2(%)	Heterogeneity	P	Egger intercept	Horizontal pleiotropy		MR-PRESSO	Steiger
			OR(95%CI)			Cochran's Q			SE	P		Correct_causal_direction
CDKN2A	EC	6	1.546 (1.298–1.842)	< 0.001	12	5.675	0.339				0.391	TRUE
CDKN2A	EC	6	1.322 (0.654–2.673)	0.48	26	5.399	0.249	0.014	0.032	0.675		
CDKN2A	EC	6	1.608 (1.276–2.025)	< 0.001								
CDKN2A	EC	6	1.678 (1.142–2.465)	0.046								
CDKN2A	EC	6	1.649 (1.137–2.392)	0.046								
DDA1	EC	19	1.111 (1.021–1.208)	0.014	0	16.632	0.549				0.607	TRUE
DDA1	EC	19	1.306 (1.042–1.635)	0.033	0	14.333	0.643	-0.02	0.013	0.148		
DDA1	EC	19	1.093 (0.970–1.230)	0.144								
DDA1	EC	19	1.112 (0.909–1.359)	0.315								
Exposure	Outcome	Nsnp	MR			Heterogeneity			Horizontal pleiotropy		MR-PRESSO	Steiger
DDA1	EC	19	1.103 (0.940–1.293)	0.245								
LRRC42	EC	18	1.112 (1.026–1.206)	0.01	0	16.78	0.469				0.491	TRUE
LRRC42	EC	18	1.165 (0.993–1.367)	0.079	2	16.331	0.43	-0.007	0.01	0.516		
LRRC42	EC	18	1.100 (0.971–1.245)	0.133								
LRRC42	EC	18	1.045 (0.846–1.290)	0.69								
LRRC42	EC	18	1.076 (0.927–1.249)	0.349								
POLB	EC	18	1.072 (1.004–1.145)	0.038	0	13.629	0.693				0.744	TRUE
POLB	EC	18	0.967 (0.864–1.084)	0.574	0	8.845	0.92	0.023	0.01	0.044		
POLB	EC	18	1.045 (0.953–1.146)	0.349								
POLB	EC	18	1.046 (0.923–1.185)	0.494								
POLB	EC	18	1.048 (0.954–1.152)	0.339								

Table 5. Results of causal association analysis of differential gene eQTLs with EC.

studies. Firstly, the prevalence of OA is high among middle-aged and older women, and a risk increase of about 10% may carry a significant disease burden at the population level. Secondly, MR analysis assess the effects of ‘lifetime exposure’ at the genetic level, so that the effect values are generally more conservative than for environmental exposures, but the causal relationships are more robust. In addition, this study further identified cofactor genes such as CDKN2A, reinforcing the biological plausibility of this risk association. Therefore, although it is only a mild risk factor, it has potential clinical translational value in disease prediction, mechanism research and management of high-risk populations. So we can consider OA a high risk factor for EC. In addition, this study has found: The common causative genes for OA and early EC—CDKN2A, DDA1, LRRC42, POLB, ADCYAP1R1C, DNMT3A, and GLRX5—along with CDKN2A, DDA1, LRRC42, and POLB specifically, are potential drug targets for EC and have diagnostic value.

OA is recognized as a systemic inflammatory chronic disease, often associated with increased levels of various inflammatory serum markers, including interleukins, adipokines, chemokines, and tumor necrosis factor- α (TNF- α). Local and systemic inflammatory responses contribute to the destruction and remodeling of articular cartilage, facilitating OA development⁷. Gamma-interferon-induced mononuclear factor (MIG/CXCL9) is a high-risk factor for EC development, as a chronic inflammatory environment can promote tumor transformation. Alterations in inflammatory markers in OA patients have been strongly linked to the development of cancers such as breast, lung, ovarian, and bladder cancers^{8,11,25}. This study clarified the causal relationship between OA and EC from a genetic perspective through MR analysis, indicating that OA is a significant risk factor for EC. Further research is needed to determine whether female patients with OA should be routinely screened for EC.

CDKN2A is a cell cycle protein-dependent kinase inhibitor that plays an important role in cell cycle regulation. Its expression products include two proteins: p16INK4a and p14ARF. It has been shown that CDKN2A can further influence colorectal tumour development and progression by regulating copper ion concentration²⁶, and copper deficiency may compromise cartilage integrity and increase the prevalence of OA²⁷. A MR study showed that physiologically higher copper circulatory status was positively associated with the risk of developing OA²⁸. Other studies have shown that abnormalities in copper metabolism also influence the development of many gynaecological tumours, including EC, and that copper may contribute to cell proliferation and neoplasia by affecting physiological processes such as mitochondrial respiration, redox, autophagy and antioxidant defences. It also promotes the growth and movement of vascular endothelial cells by regulating the synthesis and secretion of major pro-angiogenic mediators. It can also be involved in tumour spreading in conjunction with its binding proteins^{29,30}. In addition, with age, senescent cells accumulate in the body, and when joint damage occurs, senescent cells induce CDKN2A gene expression, which promotes the expression of the downstream protein p16 INK 4a, leading to synovial inflammation and accelerating the progression of OA³¹, at the same time methylation of this gene leads to p16INK4a loss of function and promotes the onset and progression of EC³². CDKN2A is involved in cell mitosis and regulates the cell cycle, with high expression correlating with poor cancer outcomes. Tumors with elevated CDKN2A expression tend to be more aggressive and have shorter

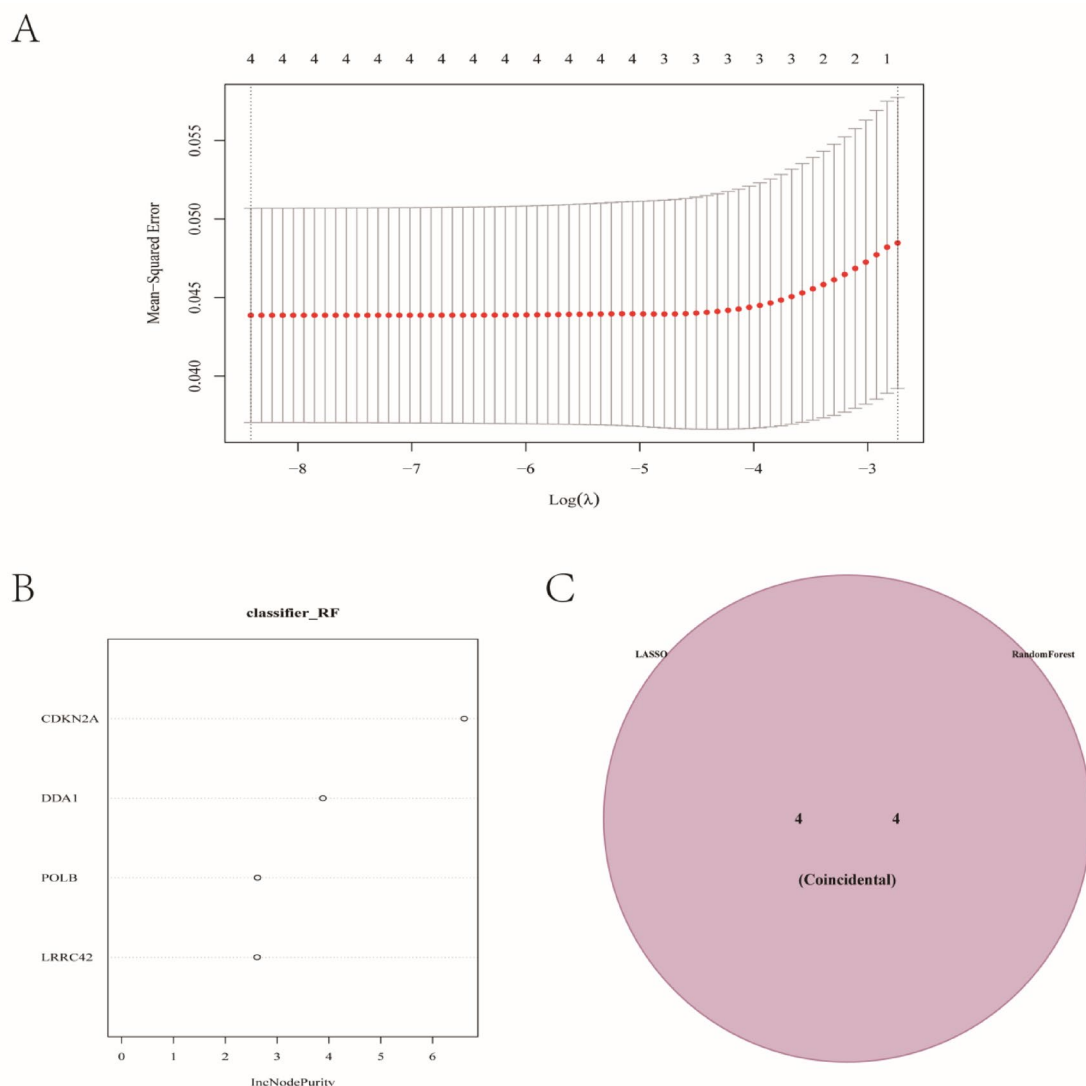


Fig. 5. Screening of genes with diagnostic value using multiple machine learning algorithms. (A: Construction of an EC prediction model using LASSO modelling B: Construction of an EC prediction model using RF modelling, C: The two models A and B take the intersection.).

recurrence intervals³³. This study demonstrated that the CDKN2A gene is a common causative gene for OA and EC, which may promote OA by affecting cellular copper metabolism, while interfering with the normal cell cycle and promoting the onset and progression of EC. Therefore, we suggest that CDKN2A may be a potential target for pharmacological intervention in OA and EC and has diagnostic value for early EC and OA.

DDA1 is a DNA damage repair-related gene that can form a complex with DET 1 and DDB 1, which functions in conjunction with ubiquitin-conjugating enzyme E2³⁴. DDA1 is also part of the core subunit of the ubiquitin ligase E3 and can control transcription-coupled repair by regulating ubiquitination activity³⁵. It has been shown that ubiquitin ligase E3 is involved in mitochondrial dysfunction, inflammatory vesicle induction, and matrix-degrading enzyme overexpression in the development of OA³⁶. Curcumin reduces inflammation and oxidative stress in OA by regulating the function of ubiquitin ligase E3³⁷. Deubiquitinating enzymes can affect chondrocyte function in OA patients by regulating the NF-κB signalling pathway and the Wnt/β-catenin signalling pathway³⁸. In addition, overexpression of DDA1 is closely associated with the development and progression of a variety of malignant tumours. DDA1 overexpression promotes lung tumour cell proliferation³⁴. In breast cancer tissues, DDA1 is a target of STAT, and its overexpression favours cancer cell proliferation metastasis and invasion. Dihydroartemisinin inhibits proliferation and induces apoptosis in cisplatin-resistant breast cancer cells by regulating the STAT3/DDA1 signalling pathway³⁹. In colon cancer tissues, DDA1 is activated, potentially promoting colon carcinogenesis through the NFκB/CSN2/GSK3β signaling pathway⁴⁰. Ubiquitin-conjugating enzyme E2 is strongly associated with EC progression, and it may have a role in promoting tumour metastasis and invasion⁴¹. DDA1 has been identified as a prevalent pathogenic gene associated with OA and EC, potentially through its influence on the regulation of ubiquitinating enzymes, which subsequently impacts the initiation and progression of both OA and early EC. Therefore, DDA1 may be a biomarker for predicting EC, and

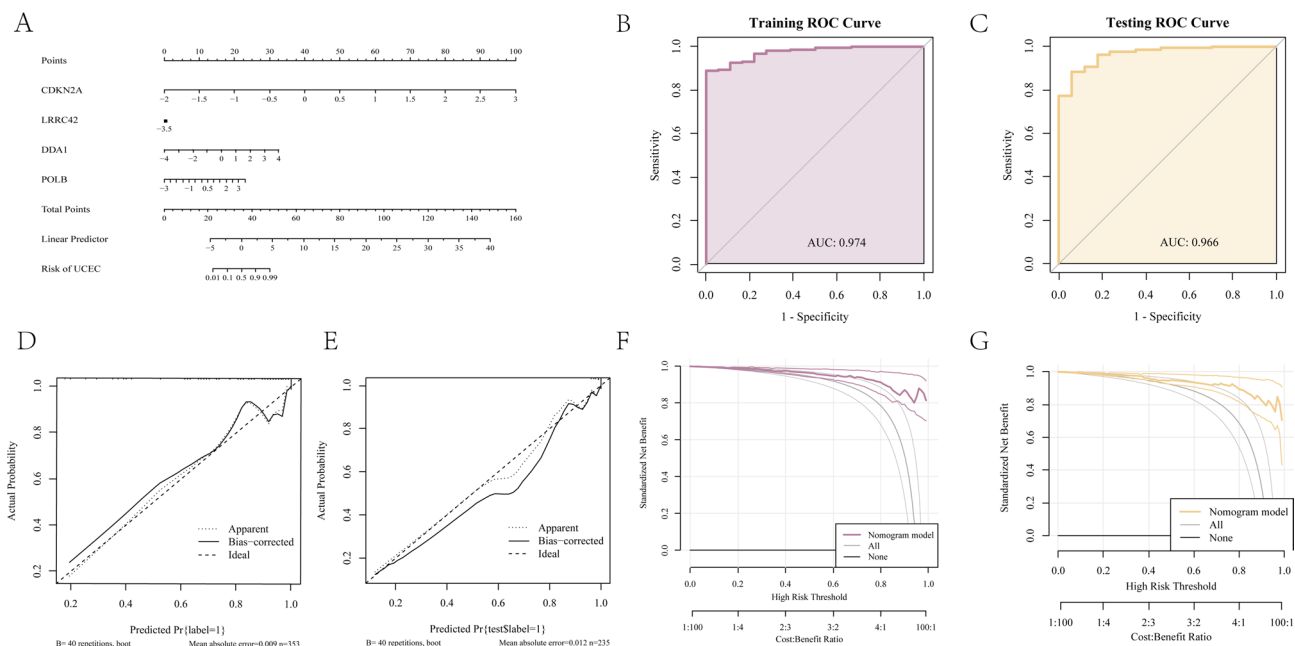


Fig. 6. Construction of the prediction model. (A) Nomogram for the 4 genes with diagnostic value (B) ROC training curves for the 4 genes with diagnostic value (C) ROC testing curves for the 4 genes with diagnostic value (D) Training curves for nomogram prediction models (E) Testing curves for nomogram prediction models (F) DCA training Curve for nomogram prediction models (G) DCA testing Curve for nomogram prediction models.

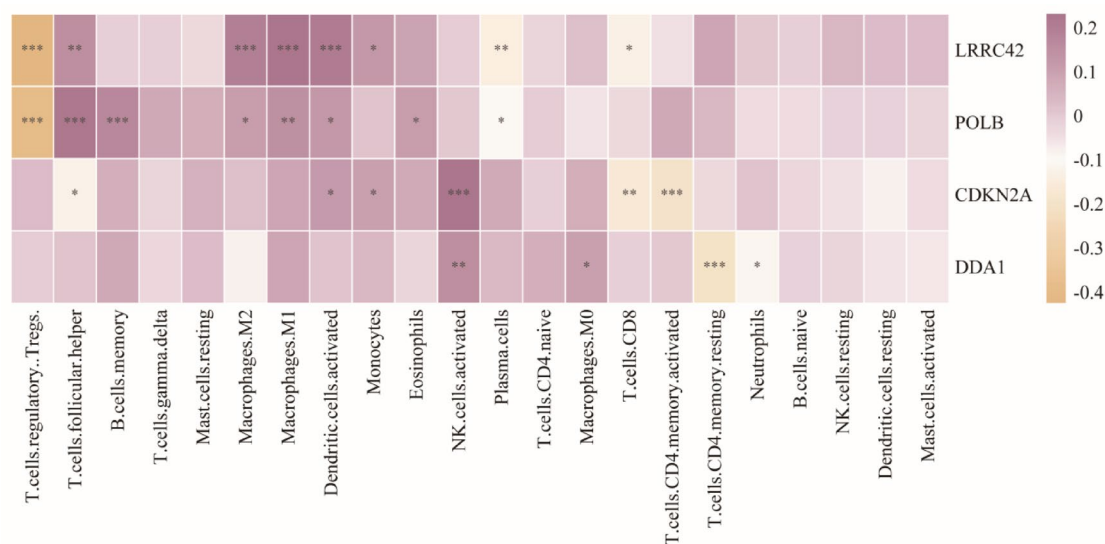


Fig. 7. Correlation analysis of immune cell infiltration (ns, $p \geq 0.05$; *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$).

Dihydroartemisinin may inhibit the progression of EC by regulating DDA1, but its specific mechanism needs to be further explored.

LRRC42 is a member of the LRR superfamily and encodes a protein characterised by leucine-rich repeat sequences. LRR proteins have multiple functions, including apoptosis, nuclear mRNA transport, cell adhesion, neuronal development, and immune response⁴². In addition, LRR superfamily expression is upregulated in many types of cancer. It has been shown that LRRC59 overexpression is associated with poor prognosis in lung cancer and promotes the proliferation and metastasis of lung cancer cells⁴³. LRRC15 plays an important role in ovarian cancer metastasis and can increase the probability of adhesion, colonisation to the omentum and invasion, suggesting that LRRC15-targeted therapy can inhibit ovarian cancer progression to a certain extent⁴⁴. However, there are relatively few reports on the role of LRRC42 in cancer progression. A study showed that

LRRC42 was highly overexpressed in lung cancer cells, and down-regulation of LRRC42 expression inhibited the growth of lung cancer cells, suggesting that LRRC42 may be an important growth-promoting factor in lung cancer cells⁴⁵. Another study showed that LRRC42 expression was increased in hepatocellular carcinoma cell lines, however, there was a significant inhibition of cell proliferation when the gene was knocked out⁴². Then, we can speculate that the occurrence of early EC may also be associated with the overexpression of LRRC42. Although the relationship between LRRC42 and OA has not been clearly reported, it has been shown that LRRC39, which is of the same family as LRRC42, is specifically highly expressed in skeletal muscle⁴⁶. LRRC42 is likely to be involved in the onset and development of OA as well. Inflammatory infiltration of the intra-articular microenvironment in patients with OA is predominantly macrophage-dominated and has been implicated as a cause of OA^{47,48}. Our study showed that LRRC42 is a common pathogenic gene in OA and EC, and its expression is significantly and positively correlated with immune cell infiltration, especially the activation status of macrophage M1, M2, and dendritic cells, which in EC suppresses the immune response through the secretion of anti-inflammatory factors, and accelerates the progression of EC through the secretion of pro-angiogenic factors that promote angiogenesis⁴⁹. In summary, LRRC42 may be mediating the onset and progression of early EC as well as OA through Macrophages M1 and Macrophages M2, and LRRC42 is expected to be a potential target for pharmacological intervention, providing new insights into the diagnosis and treatment of both diseases.

POLB is a DNA repair polymerase involved in base excision repair, recombination and drug resistance. It has been shown that colorectal cancer patients are often accompanied by overexpression of POLB⁵⁰. Overexpression of POLB leads to cisplatin resistance and poorer prognosis in colorectal cancer patients⁵¹. In gastric cancer patients, overexpression of POLB stimulates tumour proliferation and promotes invasion and metastasis⁵². Another study reinforces the important role of POLB in tumour development⁵³. In acute lymphoblastic leukaemia, POLB overexpressing tumour cells are more resistant, POLB inhibitor oleanolic acid (OA) increases the sensitivity of resistant cells to thiopurines⁵⁴. Additionally, the dRP cleavage enzyme-deficient variant of POLB (Leu22-Pro or L22P) increases genomic instability associated with mitotic dysfunction, leading to cytoplasmic DNA-mediated inflammatory responses. Inhibition of poly ADP ribose polymerase 1 exacerbates chromosomal instability and enhances cytoplasmic DNA-mediated inflammatory responses⁵⁵. In contrast, OA pathogenesis is primarily associated with a local inflammatory response⁸. Overall, the POLB gene may be linked to the development of early EC and OA, as supported by MR results from this study, though its exact pathogenesis requires further investigation.

Although our study has identified robust causal relationships using MR analyses, several factors warrant consideration. First, the study population was limited to a European cohort. The findings of the research should be approached with caution when extrapolating to other groups characterized by heightened genetic and environmental diversity (for instance, Asian populations). Second, the database is derived from public sources, features a relatively small sample size, and lacks long-term follow-up data. Finally, completely excluding the effect of potential pleiotropy remains challenging in MR studies and thorough experimental validation is essential before any clinical applications can proceed. Future studies should repeat the validation in a multiracial context to test the robustness and applicability of this study's findings.

Conclusion

OA is a significant risk factor for the development of EC and may influence its progression. The DKN2A, DDA1, LRRC42, and POLB genes could be common causative factors for both early EC and OA. This suggests that OA and EC may share a common genetic susceptibility in certain populations. Validation of this finding in a large prospective cohort could provide a basis for clinical development of screening strategies for high-risk populations. For example, regular endometrial cancer screening is recommended for female OA patients carrying high-risk variants. In addition, co-causal genes such as CDKN2A and DDA1 may become common drug intervention targets, and subsequent studies may further combine functional experiments and drug databases to explore the druggability of these genes and their potential in combination therapy.

Data availability

The data analysed in this study are available from the GEO (<http://www.ncbi.nlm.nih.gov/geo/>) (accession number: GSE12021) and TCGA (<https://portal.gdc.cancer.gov/>) databases (accession number: TCGA- UCEC), the IEU OpenGWAS project (<https://gwas.mrcieu.ac.uk/>) (accession number: OA: ebi-a-GCST007092, endometrial cancer: ebi-a-GCST006464, 7 genes: eqtl-a-ENSG00000070501, eqtl-a- ENSG00000116212, eqtl-a-ENSG00000119772, eqtl-a-ENSG00000147889, eqtl-a-ENSG00000182512, eqtl-a-ENSG00000130311) were obtained.

Received: 27 October 2024; Accepted: 27 May 2025

Published online: 01 July 2025

References

1. Crosbie, E. J. et al. Endometrial cancer. *Lancet* **399**, 1412–1428 (2022).
2. Brooks, R. A. et al. Current recommendations and recent progress in endometrial cancer. *CA Cancer J. Clin.* **69**, 258–279 (2019).
3. Shen, Y., Yang, W., Liu, J. & Zhang, Y. Minimally invasive approaches for the early detection of endometrial cancer. *Mol. Cancer* **22**, 53 (2023).
4. Huvila, J., Pors, J., Thompson, E. F. & Gilks, C. B. Endometrial carcinoma: Molecular subtypes, precursors and the role of pathology in early diagnosis. *J. Pathol.* **253**, 355–365 (2021).
5. Key, T. J. & Pike, M. C. The dose-effect relationship between 'unopposed' oestrogens and endometrial mitotic rate: Its central role in explaining and predicting endometrial cancer risk. *Br. J. Cancer* **57**, 205–212 (1988).
6. Nedunchezhiyan, U. et al. Obesity, inflammation, and immune system in osteoarthritis. *Front. Immunol.* **13**, 907750 (2022).
7. Martel-Pelletier, J. et al. Osteoarthritis. *Nat. Rev. Dis. Prim.* **2**, 12076 (2016).

8. Bouras, E. et al. Circulating inflammatory cytokines and risk of five cancers: A Mendelian randomization analysis. *BMC Med.* **20**, 1–15 (2022).
9. Pang, H. et al. Low back pain and osteoarthritis pain: A perspective of estrogen. *Bone Res.* **11**, 42 (2023).
10. Hazelwood, E. et al. Identifying molecular mediators of the relationship between body mass index and endometrial cancer risk: A Mendelian randomization analysis. *BMC Med.* **20**, 125 (2022).
11. Molnar, V. et al. Cytokines and chemokines involved in osteoarthritis pathogenesis. *Int. J. Mol. Sci.* **22**, 9208 (2021).
12. von Gruenigen, V. E. et al. Lifestyle challenges in endometrial cancer survivorship. *Obstet. Gynecol.* **117**, 93–100 (2011).
13. Lovegrove, C. E., Howles, S. A., Furniss, D. & Holmes, M. V. Causal inference in health and disease: A review of the principles and applications of Mendelian randomization. *J. Bone Miner. Res.* **39**, 1539–1552 (2024).
14. Ference, B. A. Interpreting the clinical implications of drug-target Mendelian randomization studies. *J. Am. Coll. Cardiol.* **80**, 663–665 (2022).
15. Lin, J., Zhou, J. & Xu, Y. Potential drug targets for multiple sclerosis identified through Mendelian randomization analysis. *Brain* **146**, 3364–3372 (2023).
16. Cao, Y., Yang, Y., Hu, Q. & Wei, G. Identification of potential drug targets for rheumatoid arthritis from genetic insights: A Mendelian randomization study. *J. Transl. Med.* **21**, 616 (2023).
17. Ference, B. A., Majeed, F., Penumetcha, R., Flack, J. M. & Brook, R. D. Effect of naturally random allocation to lower low-density lipoprotein cholesterol on the risk of coronary heart disease mediated by polymorphisms in NPC1L1, HMGCR, or both. *J. Am. Coll. Cardiol.* **65**, 1552–1561 (2015).
18. Swerdlow, D. I. et al. HMG-coenzyme A reductase inhibition, type 2 diabetes, and bodyweight: Evidence from genetic analysis and randomised trials. *Lancet* **385**, 351–361 (2015).
19. Schwartz, R., Gagliano Taliun, S. A. & Evans, D. M. Ten simple rules for conducting a Mendelian randomization study. *PLOS Comput. Biol.* **17**, e1009238 (2021).
20. Skrivanekova, V. W. et al. Strengthening the reporting of observational studies in epidemiology using Mendelian randomization. *JAMA* **326**, 1614–1621 (2021).
21. Ritchie, M. E. et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47–e47 (2015).
22. Kanehisa, M., Furumichi, M., Sato, Y., Matsuura, Y. & Ishiguro-Watanabe, M. KEGG: Biological systems database as a model of the real world. *Nucleic Acids Res.* **53**, D672–D677 (2025).
23. Kanehisa, M. Toward understanding the origin and evolution of cellular organisms. *Protein Sci.* **28**, 1947–1951 (2019).
24. Goto, M. K. S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **1**(28), 27–30 (2000).
25. Zhang, X. et al. The causal relationship between osteoarthritis and bladder cancer: A Mendelian randomization study. *Cancer Med.* **13**, e6829 (2023).
26. Cheng, X. et al. The crosstalk role of CDKN2A between tumor progression and cuproptosis resistance in colorectal cancer. *Aging* **16**, 10512–10538 (2024).
27. Chang, B., Hu, Z., Chen, L., Jin, Z. & Yang, Y. Development and validation of cuproptosis-related genes in synovitis during osteoarthritis progress. *Front. Immunol.* **14**, 1090596 (2023).
28. Zhou, J. et al. Genetically predicted circulating levels of copper and zinc are associated with osteoarthritis but not with rheumatoid arthritis. *Osteoarthritis Cartil.* **29**, 1029–1035 (2021).
29. Conforti, R. A., Delsouc, M. B., Zorychta, E., Telleria, C. M. & Casais, M. Copper in gynecological diseases. *Int. J. Mol. Sci.* **24**, 17578 (2023).
30. Huang, X., Lian, M. & Li, C. Copper homeostasis and cuproptosis in gynecological cancers. *Front. Cell Dev. Biol.* **12**, 1459183 (2024).
31. Jeon, O. H. et al. Local clearance of senescent cells attenuates the development of post-traumatic osteoarthritis and creates a pro-regenerative environment. *Nat. Med.* **23**, 775–781 (2017).
32. Su, L., Wang, H., Miao, J. & Liang, Y. Clinicopathological significance and potential drug target of CDKN2A/p16 in endometrial carcinoma. *Sci. Rep.* **5**, 13238 (2015).
33. Wang, J. Z. et al. Increased mRNA expression of CDKN2A is a transcriptomic marker of clinically aggressive meningiomas. *Acta Neuropathol.* **146**, 145–162 (2023).
34. Cheng, L. et al. DDA1, a novel oncogene, promotes lung cancer progression through regulation of cell cycle. *J. Cell Mol. Med.* **21**, 1532–1544 (2017).
35. Llerena, D. A. et al. The small CRL4^{CSA} ubiquitin ligase component DDA1 regulates transcription-coupled repair dynamics. *Nat. Commun.* **15**, 6374 (2024).
36. Liu, Y., Molchanov, V. & Yang, T. Enzymatic machinery of ubiquitin and ubiquitin-like modification systems in chondrocyte homeostasis and osteoarthritis. *Curr. Rheumatol. Rep.* **23**, 62 (2021).
37. Lin, Y. et al. Novel insights into the role of ubiquitination in osteoarthritis. *Int. Immunopharmacol.* **132**, 112026 (2024).
38. Wong, P. F. & Kamarul, T. Targeting Ubiquitin-Proteasome system (UPS) in treating osteoarthritis. *Eur. J. Pharmacol.* **989**, 177237 (2025).
39. Zhang, J. et al. Dihydroartemisinin affects STAT3/DDA1 signaling pathway and reverses breast cancer resistance to cisplatin. *Am. J. Chin. Med.* **51**, 445–459 (2023).
40. Zhao, S. et al. DDA1 promotes stage IIB–IIC colon cancer progression by activating NFκB/CSN2/GSK-3β signaling. *Oncotarget* **7**, 19794–19812 (2016).
41. Ma, S., Chen, Q., Li, X., Fu, J. & Zhao, L. UBE2C serves as a prognosis biomarker of uterine corpus endometrial carcinoma via promoting tumor migration and invasion. *Sci. Rep.* **13**, 16899 (2023).
42. Huang, R., Yao, J., Liu, L., Li, H. & Huang, B. Comprehensive analysis of LRRC42 as a potential biomarker and key cellular processes in cancer development. *Sci. Rep.* **15**, 5169 (2025).
43. Li, D., Xing, Y., Tian, T., Guo, Y. & Qian, J. Overexpression of LRRC59 is associated with poor prognosis and promotes cell proliferation and invasion in lung adenocarcinoma. *Onco Targets Ther.* **13**, 6453–6463 (2020).
44. Ray, U. et al. Targeting LRRC15 inhibits metastatic dissemination of ovarian cancer. *Cancer Res.* **82**, 1038–1054 (2022).
45. Fujitomo, T., Daigo, Y., Matsuda, K., Ueda, K. & Nakamura, Y. Identification of a nuclear protein, LRRC42, involved in lung carcinogenesis. *Int. J. Oncol.* **45**, 147–156 (2014).
46. Kenneth, C., Ehrlich, M. L. & Ehrlich, M. Epigenetics of skeletal muscle-associated genes in the ASB, LRRC, TMEM, and OSBP gene families. *Epigenomes* **4**, 1 (2020).
47. Wang, L. & He, C. Nrf2-mediated anti-inflammatory polarization of macrophages as therapeutic targets for osteoarthritis. *Front. Immunol.* **13**, 967193 (2022).
48. Zhang, Y. et al. Macrophage migration inhibitory factor activates the inflammatory response in joint capsule fibroblasts following post-traumatic joint contracture. *Aging* **17**(13), 5804–5823 (2021).
49. Cassetta, L. et al. Human tumor-associated macrophage and monocyte transcriptional landscapes reveal cancer-specific reprogramming, biomarkers, and therapeutic targets. *Cancer Cell* **35**, 588–602.e510 (2019).
50. Donigan, K. A. et al. Human POLB gene is mutated in high percentage of colorectal tumors. *J. Biol. Chem.* **287**, 23830–23839 (2012).
51. Iwatsuki, M. et al. A platinum agent resistance gene, POLB, is a prognostic indicator in colorectal cancer. *J. Surg. Oncol.* **100**, 261–266 (2009).

52. Tan, X. et al. Clinical significance of a point mutation in DNA polymerase beta (POLB) gene in gastric cancer. *Int. J. Biol. Sci.* **11**, 144–155 (2015).
53. Alkhanbouli, R. et al. Analysis of cancer-associated mutations of POLB using machine learning and bioinformatics. *IEEE ACM Trans. Comput. Biol. Bioinform.* **21**, 1436–1444 (2024).
54. Teng, J. Y. et al. Targeting DNA polymerase beta elicits synthetic lethality with mismatch repair deficiency in acute lymphoblastic leukemia. *Leukemia* **37**, 1204–1215 (2023).
55. Zhao, S., Goewey Ruiz, J. A., Sebastian, M. & Kidane, D. Defective DNA polymerase beta invoke a cytosolic DNA mediated inflammatory response. *Front. Immunol.* **13**, 1039009 (2022).

Author contributions

Yy.B., D.L. and S.L. conceived the project. Yy.B., S.L., Rz.S., Lw.Y. and Xm.Z. wrote the manuscript. Yy.B., S.L. and Lw.Y. performed the computational analysis. All authors read and approved the final manuscript.

Funding

This study was supported by the Yinchuan Science and Technology Tackling Project (2023SFZD02) and Special Funds for the Central Government to Guide Local Scientific and Technological Development (2024FRD05067). The funders had no roles in study design, data collection and analysis, publication decision, or manuscript preparation.

Declarations

Competing interests

The authors declare no competing interests.

Ethics approval and consent to participate

This study's data came from a European population via the publicly available GWAS database. Informed consent was obtained from the participants in the original study, which meant that ethics committee approval was not required for this aspect of the research.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-025-04470-x>.

Correspondence and requests for materials should be addressed to D.L.

Reprints and permissions information is available at www.nature.com/reprints.

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

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025, corrected publication 2025