Original Article

Exploring potential causal genetic variants and genes for endometrial cancer: Open Targets Genetics, Mendelian randomization, and multi-tissue transcriptome-wide association analysis

Guorui Zhang[1], Su Mao[1], Guangwei Yuan[2], Yang Wang[3], Jingyun Yang[4,5], Yuxin Dai[1]

[1]Department of Obstetrics and Gynecology, State Key Laboratory of Complex, Severe and Rare Diseases, National Clinical Research Center for Obstetric & Gynecologic Diseases, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China; [2]Department of Biostatistics & Bioinformatics, Roswell Park Comprehensive Cancer Center, Buffalo, NY, USA; [3]Division of Biomedical Science, Analytics and Technology, Sanofi, Toronto, ON, Canada; [4]Rush Alzheimer's Disease Center, Rush University Medical Center, Chicago, IL, USA; [5]Department of Neurological Sciences, Rush University Medical Center, Chicago, IL, USA

Contributions: (I) Conception and design: Y Dai, J Yang; (II) Administrative support: None; (III) Provision of study materials or patients: None; (IV) Collection and assembly of data: G Zhang, G Yuan, Y Wang; (V) Data analysis and interpretation: G Zhang, G Yuan, Y Wang; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Yuxin Dai, MD. Department of Obstetrics and Gynecology, State Key Laboratory of Complex, Severe and Rare Diseases, National Clinical Research Center for Obstetric & Gynecologic Diseases, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, No. 1 Shuaifuyuan Road, Dongcheng District, Beijing 100730, China. Email: helen81918@163.com.

Background: Endometrial cancer (EC) is the most common gynecological malignancy in developed countries, with incidence rates continuing to rise globally. However, the precise mechanisms underlying EC pathogenesis remain largely unexplored. This study aims to prioritize genes associated with EC by leveraging multi-omics data through various bioinformatic methods.

Methods: We utilized the Open Targets Genetics (OTG) database to pinpoint potential causal variants and target genes for EC. To explore the pleiotropic effects of gene expression on EC, we applied the Summary-based Mendelian Randomization (SMR) using summary data from a genome-wide association study (GWAS) on EC and expression quantitative trait loci (eQTL) data from the Consortium for the Architecture of Gene Expression (CAGE). We also conducted a cross-tissue transcriptome-wide association study (TWAS) employing sparse canonical correlation analysis (sCCA). Results from the sCCA TWAS and single-tissue TWAS for 22 tissues were combined using the aggregated Cauchy association test (sCCA + ACAT) to identify genes with cis-regulated expression levels linked to EC.

Results: The OTG database recognized 15 genomic loci showing independent association with EC. Gene prioritization highlighted nine genes with relatively high locus-to-gene (L2G) scores (≥0.5), the majority of which aligned with those identified using the closest gene. Colocalization analysis identified 11 additional genes at these loci. Our SMR analysis revealed two genes, *EVI2A* and *SRP14*, exhibiting a significant pleiotropic association with EC. Cross-tissue TWAS identified 31 genes whose expression was significantly associated with EC after correction for multiple testing, with four genes (*EIF2AK4*, *EVI2A*, *EVI2B*, and *NF1*) also confirmed by gene colocalization in the OTG analysis.

Conclusions: We confirmed the involvement of *EVI2A* in the pathogenesis of EC and identified several other genes that may contribute to EC development. These findings offer new insights into the genetic mechanisms underlying EC and may inform future research and therapeutic strategies.

Keywords: Endometrial cancer (EC); expression quantitative trait loci (eQTL); summary Mendelian randomization; genome-wide association study (GWAS); transcriptome-wide association study (TWAS)

Submitted May 31, 2024. Accepted for publication Sep 29, 2024. Published online Nov 21, 2024.
doi: 10.21037/tcr-24-887
View this article at: https://dx.doi.org/10.21037/tcr-24-887

© AME Publishing Company.

Introduction

Endometrial cancer (EC) is a malignancy of the inner epithelial lining of the uterus and stands as the most common gynecological malignancy in developed countries. The incidence of EC is increasing globally (1), posing a growing challenge to public health systems. EC significantly impacts patients' quality of life (2), contributing to increasing morbidity and mortality rates (3). Despite recent advances in understanding genetic diversity and identifying key drivers of various pathogenic states, achieving enhanced therapeutic precision in EC treatment remains a formidable challenge (4). Early detection and effective treatment are crucial for improving outcomes and mitigating the disease's impact on patients' lives.

Many risk factors for EC have been identified, including advanced age, obesity, exposure to radiation, and infertility, particularly in the presence of polycystic ovarian syndrome (5). EC exhibits a heterogeneous pathophysiology, with genetics playing a pivotal role in predisposition and pathogenesis. A family history of EC increases the risk by 2–3 times (6). Positive genetic correlations have been observed between EC and traits such as type 2 diabetes, body mass index (BMI), and related anthropometric characteristics, while negative correlations exist with age at menarche and years of schooling (7). These findings suggest that shared genetic backgrounds influence traits related to obesity or genetically linked to BMI, with BMI demonstrating a causal effect on EC risk in Mendelian randomization analysis (8).

Candidate gene studies have identified modest-risk variants in genes such as *ESR1*, *TERT*, *CLPTM1L*, *CYP19A1*, and *HNF1B* (9-12). Genome-wide association studies (GWASs) have pinpointed common genetic variants (minor allele frequency >1%) in about 20 potential risk loci, including regions near the *HNF1B*, *CYP19A1*, and *SOX4* genes (7,13). Functional analysis of GWAS loci revealed non-coding regions enriched for EC risk variants (7), with locus-specific studies highlighting *KLF5* and *HNF1B* as crucial susceptibility genes (12,14).

Additionally, EC was reported to be associated with several mutations that vary across its subtypes, involving genes such as *PTEN*, *PIK3CA/PIK3R1*, *CTNNB1*, *ARID1A*, *K-RAS* as well as *BRCA1/2* and Lynch syndrome genes like *MLH1*, *MSH2*, *MSH6*, and *PMS2* (15-20). These genes are implicated in various molecular functions, such as tumor suppression, cell proliferation and differentiation, chromatin remodeling, and DNA mismatch repair. Transcription analysis has further revealed potential key genes implicated in EC. For instance, research on early-stage EC identified over 900 differentially expressed transcripts, with four genes validated by quantitative polymerase chain reaction (PCR), including *RORB*, *IHH*, *DLGAP5*, and *MELK* (21). Another study utilizing system bioinformatics analysis identified four genes related to EC: *TOP2A* and *ASPM* were upregulated, while *EFEMP1* and *FOXL2* were downregulated in EC tissues or cells (22). Single-cell transcriptomic analysis has uncovered oncogenic subpopulation signature genes that contribute to the pathological processes in endometrial carcinoma, providing deep insights into tumor heterogeneity (23). Additionally, analysis of transcription factor binding regulatory patterns has shown an association between specific genes with EC development, indicating that transcription factor binding site analysis is effective in screening for cancer-associated genes (24).

Despite these advances, it is estimated that over 1,000

Highlight box

Key findings

- We confirmed the involvement of EVI2A in the pathogenesis of endometrial cancer (EC) and identified additional genes that may contribute to the etiology of EC.

What is known and what is new?

- EC has been extensively studied through genome-wide association studies (GWAS), which have identified numerous risk loci and candidate genes. Previous research has highlighted essential genes implicated in various pathways of EC pathogenesis. However, existing studies often fall short in systematically prioritizing causal variants and integrating functional genomics data, which are crucial for understanding the biological mechanisms driving EC.
- Our study builds upon the foundational work of earlier GWAS by leveraging the Open Targets Genetics database to identify and prioritize causal genetic variants for EC systematically. Additionally, we employed Summary-based Mendelian Randomization (SMR) and cross-tissue transcriptome-wide association studies (TWAS) to explore pleiotropic associations and cis-regulated gene expression linked to EC risk. These approaches offer new insights into the genetic underpinnings of EC that were not fully captured in previous studies.

What is the implication, and what should change now?

- Our findings suggest that integrating multi-omics data can enhance our understanding of the genetic mechanisms underlying EC. This approach may improve risk stratification and inform the development of targeted therapies.
- Further exploration of the functions of the identified genes is necessary, as they hold the potential to refine treatment strategies and improve patient outcomes.

© AME Publishing Company.

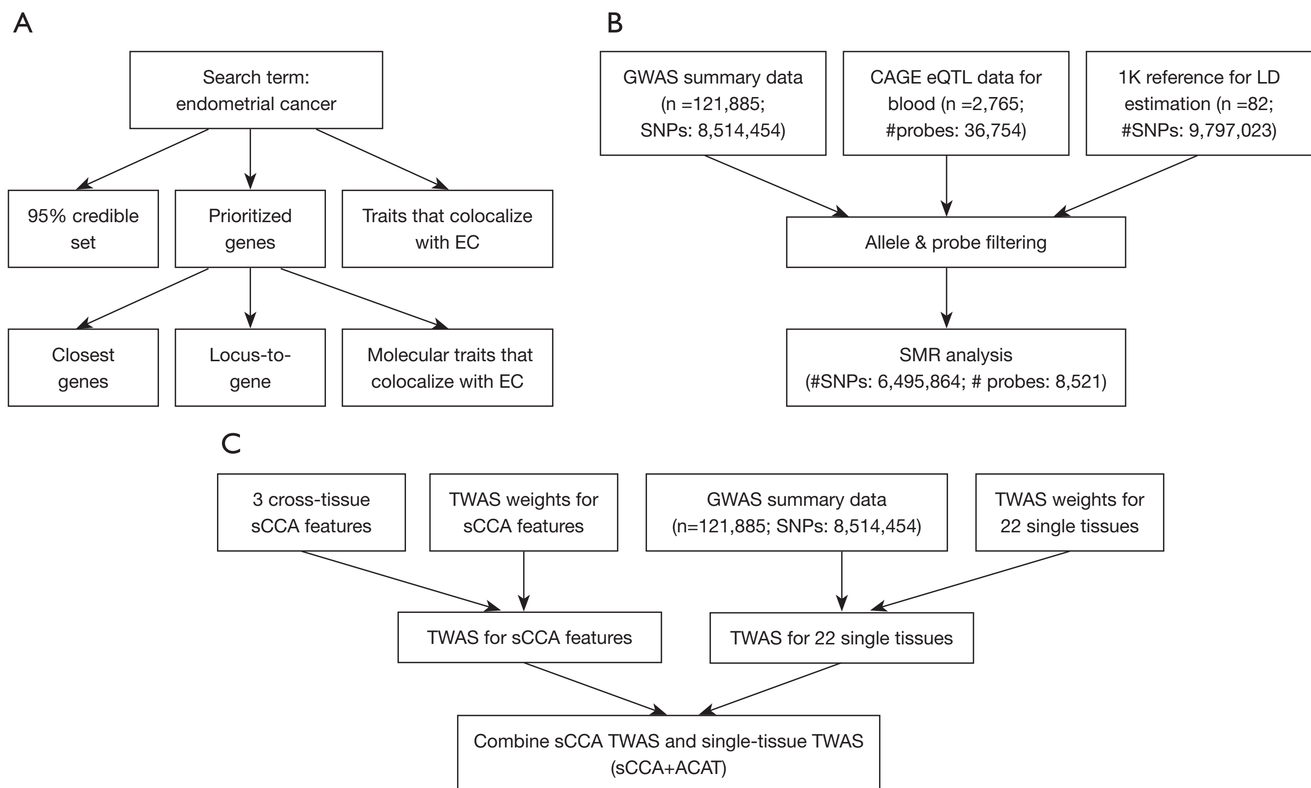

Figure 1 Flow chart for the bioinformatic analyses. (A) OTG analysis; (B) SMR analysis using CAGE eQTL data for blood; and (C) cross-tissue TWAS. ACAT, aggregated Cauchy association test; CAGE, Consortium for the Architecture of Gene Expression; eQTL, expression quantitative trait loci; LD, linkage disequilibrium; OTG, Open Targets Genetics; sCCA, sparse canonical correlation analysis; SMR, Summary-based Mendelian Randomization; SNP, single nucleotide polymorphism; TWAS, transcriptome-wide association study; EC, endometrial cancer; GWAS, genome-wide association study.

independent risk loci exist for EC (25), with only a small fraction identified to date. While GWASs have significantly advanced our understanding of EC's genetic basis, they often fall short in systematically prioritizing causal variants and elucidating their functional roles. The scarcity of integration between GWAS and functional genomics data further limits our understanding of EC's genetic architecture. Our study addresses these gaps by leveraging the Open Targets Genetics (OTG) database to identify and prioritize causal genetic variants (26). Additionally, we utilize Summary-based Mendelian Randomization (SMR) to explore pleiotropic associations (27) and cross-tissue transcriptome-wide association studies (TWASs) (28) to link cis-regulated gene expression to EC risk. These integrative approaches aim to enhance our understanding of EC's genetic mechanisms, potentially informing targeted therapeutic strategies. We present this article in accordance with the STROBE-MR reporting checklist (available at https://tcr.amegroups.com/article/view/10.21037/tcr-24-887/rc).

Methods

Editorial policies and ethical considerations

This study utilized publicly available GWAS summary results for EC and expression quantitative trait loci (eQTL) data. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). The analytical process employed in this study is detailed in *Figure 1*.

Data sources

GWAS data for EC

The GWAS summary data for EC were sourced from a recent genome-wide association meta-analysis, which included 17 studies identified through the Endometrial

© AME Publishing Company.

Cancer Association Consortium (ECAC), the Epidemiology of Endometrial Cancer Consortium, and the UK biobank (7). The meta-analysis encompassed a total of 121,885 participants of European ancestry, including 12,906 EC cases and 108,979 country-matched controls. Genotyping was done using various arrays, such as the "OncoArray" genotyping chip, the Illumina Human OmniExpress array, and the Illumina Human 660W array. Genotype data underwent quality control and imputation using the 1000 Genome Project v3 reference panel or the combined 1000 Genome Project v3 and UK10K reference panels. An additive genetic model was assumed by all participating studies, with population stratification accounted for using relevant principal components. GWAS results from individual studies were combined using a fixed-effect inverse-variance weighted meta-analysis. The GWAS summary data can be downloaded at https://www.ebi.ac.uk/gwas/studies/GCST006464.

eQTL data

We utilized the Consortium for the Architecture of Gene Expression (CAGE) eQTL summary data derived from peripheral blood samples of 2,765 participants (29). The data can be accessed at https://yanglab.westlake.edu.cn/data/SMR/cage_eqtl_data_hg19.tgz.

Refining GWAS signals and potential causal genes

To identify potential causal variants and target genes for EC, we utilized the OTG database (30) (accessed March 17, 2023). This online resource integrates GWAS and functional genomics data to systematically identify and prioritize likely causal genetic variants and genes for various traits. We searched the term 'Endometrial cancer' in OTG, which then calculated a 95% credible set for each locus independently associated with EC. The credible set represents a group of genetic variants that are 95% likely to contain the true causal variant, assuming that only one causal variant exists and has been measured. To prioritize putative causal genes from association signals, OTG offers several approaches: (I) closest genes, which identifies the gene closest to the transcription start site; (II) locus-to-gene (L2G) score, which ranks genes based on a machine learning algorithm trained on over 400 gold-standard positive GWAS loci out of 133,441 loci from all available GWAS studies. An L2G score ranges from 0 to 1, with 1 indicating the highest confidence in assigning a gene to a trait at a given locus; and (III) colocalization, which identifies molecular traits that colocalize with EC at a specific locus. OTG also reports other traits that colocalize with EC at a given locus based on previous GWAS studies.

Statistical analysis

SMR analysis

The Mendelian analysis was conducted using the method as implemented in the software SMR (27). SMR applies the principles of MR integrating GWAS and eQTL summary statistics to explore the pleiotropic association between gene expression and a trait. In SMR, the effect of gene expression on the trait was estimated by using the top cis-eQTL as the instrumental variable. The SMR analysis was performed following a similar approach as described in a previous publication (31), using default parameter settings (Table S1). Multiple testing was adjusted using the false discovery rate (FDR).

Cross-tissue TWAS analysis

To further investigate genes whose cis-regulated expression is associated with EC, we conducted a cross-tissue TWAS using the Functional Summary-based Imputation (FUSION) (32). This approach integrates GWAS summary statistics for EC with pre-computed gene expression weights. Unlike single-tissue TWAS, the cross-tissue TWAS leverages gene expression data from multiple tissues through sparse canonical correlation analysis (sCCA-TWAS) (28). This method enhances the power to detect trait-associated genes while controlling the type I error in the absence of an association. Specifically, three sCCA features were generated, each treated as repeated measures of gene expression across tissues. The TWAS method was applied for each sCCA feature, including only those that passed the heritability assessment. Additionally, a set of 22 tissues was chosen, and TWAS was executed for each one. Subsequently, the results of the sCCA TWAS and single-tissue TWAS were consolidated with the aid of the aggregated Cauchy association test (sCCA + ACAT) (33). Pre-computed weights using the Genotype-Tissue Expression (GTEx) V8 were utilized for both the sCCA features and the 22 tissues (34). We applied FDR to correct for multiple testing in the sCCA + ACAT results.

Data curation and statistical/bioinformatical analysis were performed using R version 4.2.3 (https://www.r-project.org/), PLINK 1.9 (https://www.cog-genomics.org/plink/1.9/), SMR (https://yanglab.westlake.edu.cn/software/smr/#Overview) and FUSION (http://gusevlab.org/projects/fusion/).

Table 1 Basic information of the eQTL and GWAS data

Data source	Total number of participants	Number of eligible genetic variants or probes
SMR		
eQTL	2,765	8,521
GWAS	121,885	6,495,864
Cross-tissue TWAS		
eQTL	101	11,389
GWAS	121,885	8,514,454
GWAS data used by Open Target Genetics	121,885	NA

The eligible numbers of genetic variants for SMR and cross-tissue TWAS differ due to filtering procedures: the number of genetic variants for SMR is the final number of SNPs that pass the initial filtering as specified in Table S1 while the number for cross-tissue TWAS represents the overall potentially eligible genetic variants in TWAS analyses in different tissues. eQTL, expression quantitative trait loci; GWAS, genome-wide association studies; SMR, Summary-based Mendelian Randomization; TWAS, transcriptome-wide association study; NA, not available.

Results

Basic information of the summarized data

In the SMR analyses, the CAGE eQTL data included 2,765 participants and 8,521 probes. Following allele frequencies checking among the datasets and LD pruning, approximately 6.5 million SNP were deemed eligible for the SMR analysis. For the multi-tissue TWAS analysis, around 8.5 million SNPs were used as the input. Detailed information is provided in *Table 1*.

Refining GWAS signals and potential causal genes

OTG identified 15 genomic loci showing independent association with EC (*Table 2*), located on chromosomes 1, 2, 6, 8, 9, 11, 12, 13, 15, and 17. The number of genetic variants in a credible set varies from 1 (locus 8 on chromosome 8) to 91 (locus 6 on chromosome 6). Gene prioritization using the L2G identified nine genes with relatively high L2G scores (≥0.5), including *BCL11A*, *SOX4*, *HEY2*, *DMRTA1*, *WT1*, *SSPN*, *SH2B3*, *TBX3*, and *CYP19A1*. Most of these genes were also identified using the closest gene. Additionally, colocalization analysis identified 11 additional genes that colocalize at these loci (*Table 2*). Several traits, such as endometrioid histology, menorrhagia, and sex hormone-binding globulin levels, were found to colocalize with EC at these loci.

Pleiotropic association with EC

Our SMR analysis identified four genes, tagged by six probes, that showed potential pleiotropic associations with EC (*Table 3*; Table S2), including *SKAP1* [ILMN_1751400, β (SE) =−0.18 (0.03), P_{SMR} =7.19×10^{-8}; *Figure 2*], *EVI2A* [ILMN_2369018, β (SE) =0.10 (0.02), P_{SMR}=2.70×10^{-6}; ILMN_1733579, β (SE) =−0.16 (0.02), P_{SMR}=1.09×10^{-5}; Figure S1], *SRP14* [ILMN_1809347, β (SE) = −0.41 (0.09), P_{SMR}=8.49×10^{-6}; Figure S2], and *SNX11* [ILMN_1683950, β (SE) =0.14 (0.03), P_{SMR}=1.80×10^{-5}; ILMN_1696051, β (SE) =0.18 (0.04), P_{SMR} =2.65×10^{-5}; *Figure 2*]. However, two genes, *SNX11* (tagged by two probes) and *SKAP1*, had low HEIDI P values, suggesting that their pleiotropic associations might result from linkage between the top associated cis-eQTL and two distinct causal variants, one affecting gene expression and the other affecting trait variation. The remaining two genes, *EVI2A* (tagged by two probes) and *SRP14*, passed the HEIDI test, indicating a true pleiotropic association. These two genes were also identified through gene prioritization using the closest gene and colonization in OTG analysis. Notably, *EVI2A*, tagged by ILMN_2369018, also withstood stringent Bonferroni correction (0.05/8,521=5.87×10^{-6}).

Cis-regulated gene expression in association with EC

The cross-tissue sCCA + ACAT analysis for EC was based on TWAS results from 22 issues (Table S3). We identified significant associations with EC (FDR <0.05) in 19 out of the 22 examined tissues (Table S3), involving 39 unique genes. Notably, *SNX11* was the most frequently associated gene, appearing in 14 tissues, followed by *EIF2AK4* in 10 tissues and *RP5-890E16.2* in 7 tissues. The multi-tissue TWAS using sCCA + ACAT revealed 31 significant genes whose expression was associated with EC after correction for multiple testing (FDR <0.05), with *TNFAIP8L3*, *HECTD4*, and *EIF2AK4* emerging as the top three genes (*Table 4*). Of these, *EVI2A* was also identified by SMR analysis and gene prioritization using the closest gene approach in the OTG analysis. Additionally, four of the genes (*EIF2AK4*, *EVI2A*, *EVI2B*, and *NF1*) were also identified by gene co-

© AME Publishing Company.

Table 2 Refining GWAS signals and potential causal genes using OTG

Locus	Lead variant	P value	Odds ratio	95% confidence interval	Credible set size	LD set size	Closest gene	L2G	Overall L2G score	Colocalized genes
1	1_37607755_T_C	3.58×10^{-8}	1.23	1.1–1.3	5	4	*GNL2*			
2	2_60670444_G_A	3.39×10^{-8}	1.26	1.2–1.4	14	24	*PAPOLG*	*BCL11A*		
3	6_21648854_G_A	4.15×10^{-16}	0.87	0.84–0.90	2	3	*SOX4*	*SOX4*		
4	6_125687226_A_G	2.91×10^{-10}	0.91	0.88–0.93	91	185	*HEY2*	*HEY2*		
5	8_128587032_C_G	3.11×10^{-12}	0.86	0.82–0.89	1	46	*MYC*			
6	9_22207038_T_C	6.38×10^{-9}	0.85	0.80–0.89	17	30	*CDKN2B*	*DMRTA1*		
7	11_32468118_C_T	1.33×10^{-8}	1.09	1.1–1.1	23	70	*WT1*	*WT1*		
8	12_26273405_G_A	1.10×10^{-9}	1.11	1.1–1.2	6	13	*BHLHE41*	*SSPN*		*BHLHE41*
9	12_111446804_T_C	1.14×10^{-10}	1.10	1.1–1.1	7	19	*SH2B3*	*SH2B3*		*HVCN1*, *TRAFD1*
10	12_114776743_C_T	3.47×10^{-9}	1.10	1.1–1.1	7	19	*TBX3*	*TBX3*		
11	13_73238004_C_T	2.70×10^{-17}	0.86	0.83–0.89	4	7	*KLF5*			
12	15_40029923_T_C	5.07×10^{-9}	1.09	1.1–1.1	34	29	*SRP14*			*EIF2AK4*, *SRP14*, *CCDC32*
13	15_51261712_A_G	3.30×10^{-14}	1.12	1.1–1.2	34	102	*CYP19A1*	*CYP19A1*		*CYP19A1*
14	17_31319014_G_A	4.29×10^{-8}	0.91	0.88–0.94	71	337	*EVI2A*			*NF1*, *EVI2A*, *EVI2B*, *RAB11FIP4*, *OMG*
15	17_48216874_C_A	4.66×10^{-9}	1.10	1.1–1.1	63	96	*SNX11*			

Odds ratio was calculated with respect to the alternative allele; Closest gene means the gene with the closest transcription start site; colocalized genes mean the genes which colocalize at this locus with PP(H4) ≥0.95 and $\log_2$(H4/H3) ≥$\log_2$(5). Credible size means the number of variants in the 95% credible set at this locus; LD set size means the number of variants in LD (R^2 ≥0.7) with this lead variant; L2G means the genes prioritized by the locus-to-gene model with score ≥0.5. GWAS, genome-wide association study; OTG, Open Target Genetics; LD, linkage disequilibrium; L2G, locus-to-gene.

Table 3 The probes showing potential pleiotropic association with EC in the SMR analysis*

Probe	Gene	CHR	Top SNP	P_{eQTL}	P_{GWAS}	Beta	SE	P_{SMR}	P_{HEIDI}	Q value
ILMN_1751400	*SKAP1*	17	rs2938483	3.32×10^{-68}	1.49×10^{-8}	−0.185	0.034	7.19×10^{-8}	1.25×10^{-5}	0.0006
ILMN_2369018	*EVI2A*	17	rs7505	2.26×10^{-151}	1.85×10^{-6}	0.102	0.022	2.70×10^{-6}	0.379	0.012
ILMN_1809347	*SRP14*	15	rs17722526	1.41×10^{-13}	2.46×10^{-8}	−0.410	0.092	8.49×10^{-6}	0.225	0.023
ILMN_1733579	*EVI2A*	17	rs2525570	3.63×10^{-58}	4.80×10^{-6}	−0.161	0.037	1.09×10^{-5}	0.425	0.023
ILMN_1683950	*SNX11*	17	rs62064953	2.15×10^{-77}	1.05×10^{-5}	0.140	0.033	1.80×10^{-5}	2.49×10^{-6}	0.031
ILMN_1696051	*SNX11*	17	rs12949879	3.09×10^{-49}	1.17×10^{-5}	0.179	0.043	2.65×10^{-5}	5.61×10^{-5}	0.038

*, the GWAS summarized data can be downloaded at https://www.ebi.ac.uk/gwas/studies/GCST006464. The CAGE eQTL data can be downloaded at https://cnsgenomics.com/data/SMR/#eQTLsummarydata. P_{eQTL} is the P value of the top associated cis-eQTL in the eQTL analysis, and P_{GWAS} is the P value for the top associated cis-eQTL in the GWAS analysis. Beta is the estimated effect size in SMR analysis, SE is the corresponding standard error, P_{SMR} is the P-value for SMR analysis and P_{HEIDI} is the P value for the HEIDI test. Q-value is the adjusted P value using FDR. CAGE, Consortium for the Architecture of Gene Expression; CHR, chromosome; EC, endometrial cancer; eQTL, expression quantitative trait loci; FDR, false discovery rate; GWAS, genome-wide association study; HEIDI, heterogeneity in dependent instruments; SE, standard error; SMR, summary data-based Mendelian randomization; SNP, single-nucleotide polymorphism.

© AME Publishing Company.

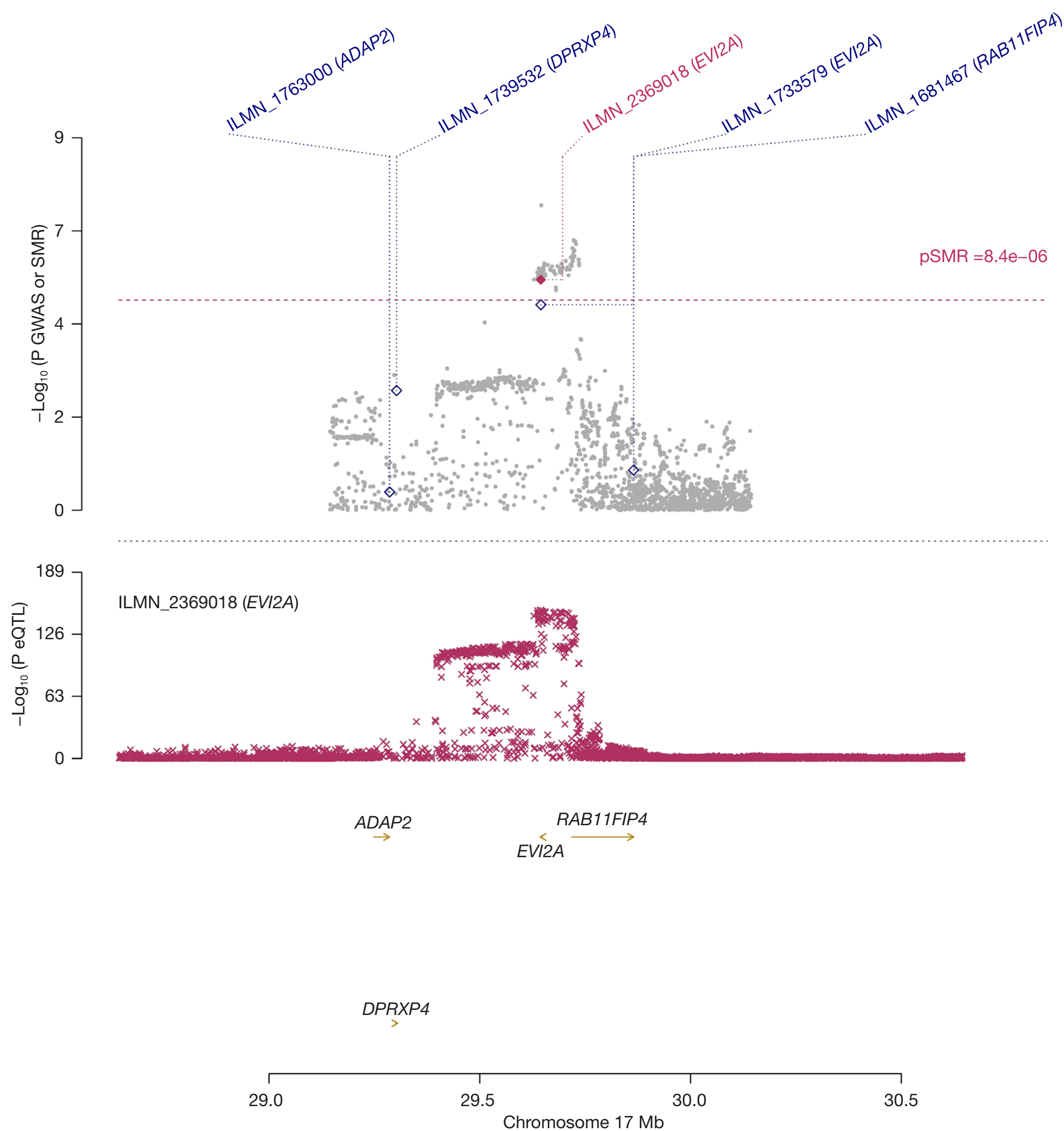

Figure 2 Pleiotropic association of *EVI2A* with EC. Top plot, grey dots represent the –lg(P values) for SNPs from the GWAS of EC, with solid rhombuses indicating that the probes pass HEIDI test. Middle plot, eQTL results. Bottom plot, location of genes tagged by the probes. EC, endometrial cancer; eQTL, expression quantitative trait loci; GWAS, genome-wide association study; HEIDI, heterogeneity in dependent instruments; SMR, Summary-based Mendelian Randomization; SNP, single nucleotide polymorphism

localization in the OTG analysis. Therefore, *EVI2A* stands out as the gene identified consistently across all three analyses.

Discussion

In this study, we sought to prioritize genes associated with EC using multiple analytical methods. *EVI2A* was consistently identified across all three approaches, confirming its role in the pathogenesis of EC. The use of different analytical tools also highlighted several other genes that may contribute to the etiology of EC, offering valuable insights into its underlying mechanisms.

A prior multi-tissue TWAS study utilized cis-eQTL

Table 4 Significant genes identified by sCCA + ACAT

Genes	Overall P	Min P	FDR
EEFSEC	2.48×10^{-6}	6.49×10^{-7}	0.004
HECTD4	4.60×10^{-8}	5.13×10^{-9}	3.65×10^{-4}
EIF2AK4	8.55×10^{-8}	1.34×10^{-8}	4.14×10^{-4}
ATF7IP2	3.15×10^{-6}	5.10×10^{-7}	0.005
EIF3CL	1.50×10^{-5}	1.74×10^{-6}	0.015
EIF3C	1.95×10^{-5}	1.92×10^{-6}	0.018
COPZ2	2.39×10^{-5}	4.22×10^{-6}	0.018
SNX11	5.02×10^{-7}	3.31×10^{-8}	0.002
LRRC37A16P	4.14×10^{-5}	5.52×10^{-6}	0.026
TEFM	2.26×10^{-6}	1.74×10^{-7}	0.004
NUPR1	2.32×10^{-5}	4.19×10^{-6}	0.018
RP5-890E16.2	1.04×10^{-7}	1.41×10^{-8}	4.14×10^{-4}
NFE2L1	1.32×10^{-6}	2.26×10^{-7}	0.003
CBX1	7.25×10^{-5}	1.51×10^{-5}	0.041
TNFAIP8L3	4.99×10^{-9}	9.98×10^{-10}	7.92×10^{-5}
ADSSL1	8.36×10^{-5}	4.05×10^{-5}	0.046
FAM46C	4.99×10^{-5}	4.99×10^{-5}	0.030
HNRNPA3P9	1.32×10^{-6}	3.31×10^{-7}	0.003
NAV3	9.20×10^{-5}	4.60×10^{-5}	0.047
BPTF	3.13×10^{-5}	1.16×10^{-5}	0.023
RUVBL1	1.25×10^{-5}	3.12×10^{-6}	0.013
NCOA7	4.06×10^{-5}	1.02×10^{-5}	0.026
SEC61A1	3.65×10^{-5}	1.63×10^{-5}	0.025
EVI2A	3.65×10^{-6}	2.23×10^{-6}	0.005
NOL11	2.20×10^{-5}	1.10×10^{-5}	0.018
C1orf74	6.00×10^{-6}	2.00×10^{-6}	0.007
PRR15L	2.12×10^{-5}	7.18×10^{-6}	0.018
EVI2B	9.59×10^{-6}	4.80×10^{-6}	0.011
NF1	4.23×10^{-6}	3.28×10^{-6}	0.006
NPIPB6	6.00×10^{-5}	2.11×10^{-5}	0.035
GTF2IRD2P1	9.11×10^{-5}	3.24×10^{-5}	0.047

Overall P is the P value from sCCA + ACAT; Min P is the minimal P value from single-tissue TWAS. ACAT, aggregated Cauchy association test; FDR, false discovery rate; sCCA, sparse canonical correlation analysis; TWAS, transcriptome-wide association study.

summary statistics and gene expression data from various tissues, including subcutaneous adipose, visceral omentum adipose, ovary, uterus, vagina, and whole blood (GTEx V8) (35). Our approach, employing sCCA TWAS + ACAT method, differs from the earlier study in several significant ways: (I) the original TWAS was based on six tissues, while our sCCA TWAS drew on TWAS results for three cross-tissue features and 22 individual tissues; (II) our sCCA TWAS + ACAT approach aggregates results using ACAT, in contrast to the S-MultiXcan and Joint-Tissue Imputation (JTI) employed by the original study. The sCCA TWAS + ACAT method has been demonstrated to provide superior statistical power (28). The significance of our study lies in several key findings. We reaffirmed the association of three genes as identified by the S-MultiXcan-based TWAS in the original study, namely *EVI2A*, *SNX11*, and *EEFSEC*. We also confirmed the association of four genes determined by the multi-tissue TWAS analysis using JTI, including *EEFSEC*, *EIF2AK4*, *SNX11*, and *NPIPB6*. Additionally, our study identified new genes potentially linked to EC, such as *TNFAIP8L3*, *HECTD4*, and *RP5-890E16.2*. These findings help enhance our understanding of EC's genetic landscape and inform the development of new therapeutic targets.

The three bioinformatic methods, namely OTG, SMR, and sCCA + ACAT, all leverage the same GWAS data on EC (7) but differ significantly in their approaches. The OTG database integrates GWAS and functional genomics data to systematically identify and prioritize likely causal genetic variants and genes. SMR, in contrast, combines GWAS and eQTL data from a single tissue (peripheral blood in our study) to infer potential causal relationships between gene expression and EC, using genetic variants as instrumental variables. Meanwhile, sCCA + ACAT analyzes gene expression across multiple tissues, enhancing the detection of gene-trait associations by incorporating data from various tissue types, thus providing a broader view of gene expression related to EC.

Our study confirms the involvement of *EVI2A* in the pathogenesis of EC through various bioinformatical approaches, aligning with the findings from previous studies (7,35). The *EVI2A* gene, also known as ecotropic viral integration site 2A, is located on chromosome 17q11.2 and is potentially associated with other proteins within a cell surface receptor complex on the membrane (36). Research suggests that *EVI2A* functions as an oncogene (37), with notable overexpression observed in oral tongue

squamous cell carcinoma (38) and osteosarcoma (OS) (39). High *EVI2A* expression was associated with worse overall survival in patients with OS, while *EVI2A* knockdown has been shown to suppress cell proliferation and migration in OS by inactivating the MEK/ERK signaling pathway (39), a pathway integral to cell proliferation, differentiation, apoptosis, migration, and cancer progression (40-42). In mice, multiple leukemogenic retroviruses integrate near the *EVI2A* gene in lymphocytes, altering its expression and suggesting a possible role as a tumor suppressor in lymphocytes (43). Moreover, initial evidence indicates that *EVI2A* may influence B cell receptor (BCR) signaling, possibly contributing to the dynamic assembly of BCR clusters (44). The gene's proximity to *NF1*, a well-known tumor suppressor, further supports the hypothesis that dysregulation of *EVI2A* could affect *NF1* function or expression, thereby influencing tumorigenesis (45). Previous research highlighted the genetic variant rs1129506 in *EVI2A* as a novel variant associated with EC (7). Additionally, two other genetic variants in *EVI2A*, rs9894648 and rs3837848, have been linked to sex hormone-binding globulin levels (SHBG) and testosterone levels (46), factors genetically associated with EC risk. Despite these findings, the precise function of *EVI2A* remains poorly understood, warranting further investigation to clarify its role in EC pathogenesis.

Our study also identified the *SNX11* gene, known as sorting nexin family member 11, through OTG and multi-tissue TWAS analysis. Located on chromosome 17q21.32, *SNX11* is part of the sorting nexin family, which contains a phox (PX) domain crucial for intracellular trafficking. Structural studies of human *SNX11* revealed a novel extended PX domain (PXe), featuring two additional α-helices beyond the conventional PX domain. This PXe is essential for inhibiting the vacuolation activity induced by *SNX10* (47), which can disrupt cellular processes, including endosomal and lysosomal trafficking, thereby affecting endosome homeostasis (48). *SNX11* interacts with various phosphoinositides, showing a strong preference for binding to phosphatidylinositol 3-phosphate (PtdIns3P), a key marker of endosomal membranes. This suggests that *SNX11* plays a vital role in endosomal trafficking and sorting, a function vital for maintaining cellular homeostasis by regulating protein movement and degradation within the cell (49). Prior research identified the genetic variant rs882380 in *SNX11* as a novel variant associated with EC (7). Further, *SNX11* has been identified as a potential target for EC risk variation through enhancer-promoter chromatin looping analysis (50). Lead credible variants (CVs) from blood eQTL data also pointed to 17q21.32, with rs882380 being among the top eQTLs (50). A phenome-wide association study (PheWAS) revealed associations between the *SNX11* and various phenotypic categories, such as cardiovascular disorders, diabetes, and sex hormones (35). However, our comprehension of *SNX11*'s precise role is limited, particularly in relation to EC. Therefore, further research is needed to elucidate the roles of *SNX11* in the pathogenesis of EC.

Our multi-tissue TWAS identified several genes previously linked to EC through cis-expression analysis. For example, the *TEFM* gene, located on chromosome 17q11.2, demonstrated a significant association with EC ($P=1.74\times10^{-7}$). *TEFM* encodes the mitochondrial transcription elongation factor, which enhances the processivity of mitochondrial RNA polymerase, POLRMT (51). Mitochondria, given their critical roles in cellular metabolism, apoptotic regulation, maintenance of redox balance, and the activation of integrated stress responses and associated immune responses (52), have become a focal point in cancer research. Previous research identified a genetic variant in *TEFM* (rs1129506) as being associated with EC [odds ratio (OR) =0.91, 95% confidence interval (CI): 0.88–0.94, $P=4.3\times10^{-8}$] (7). Additionally, elevated *TEFM* expression has been shown to promote growth and metastasis in hepatocellular carcinoma by activating reactive oxygen species (ROS) and extracellular signal-regulated kinase (ERK) signaling (53). Despite these findings, the precise role of *TEFM* in the pathogenesis of EC remains unclear, underscoring the need for further investigation.

Our study has several limitations. While we identified multiple genes showing pleiotropic association with EC, we were unable to directly compare the expression of these genes between EC patients and the control group due to the lack of relevant gene expression data. Future studies should investigate gene expression changes to better understand the potential pathogenic mechanisms involved. The incidence of EC varies across ethnicities, suggesting ethnic-specific genetic architecture. However, the GWAS summary data in our analyses were derived from participants of European ancestry, and we also lacked ethnicity-specific gene expression and eQTL data. As a result, our findings may not be generalizable to other ethnic groups, underscoring the need for further studies to compare gene expression across different populations. Previous research suggests that the SMR approach performs well with a sample size of 1,000 for eQTL summarized

© AME Publishing Company.

data and 10,000 for GWAS summarized data (27). Given this, power is not a significant concern as our study utilized CAGE eQTL data from 2,765 subjects and GWAS summarized data from 121,885 subjects. The sample size of available uterus eQTL data is limited (e.g., n=150 in GTEx V8). Therefore, we used the CAGE eQTL data from peripheral blood. Since eQTL data are tissue-specific, future studies with larger sample sizes for uterus eQTL data are needed to validate our findings. Similarly, we employed the multi-tissue TWAS approach (i.e., sCCA TWAS + ACAT) rather than a tissue-specific TWAS (e.g., uterus) due to the limited sample size of available eQTL data for the uterus. Although the sCCA TWAS + ACAT method offers substantially higher power than traditional single-tissue TWAS methods in identifying genes with genetically predicted expression associated with a trait, it may obscure genetic associations that are tissue-specific, which may be crucial for understanding the specific pathways involved in EC and for developing targeted prevention and treatment strategies. The multi-tissue TWAS provided only tissue-specific test statistics (Table S3) without the overall effect size and the corresponding direction due to its cross-tissue nature. The number of eligible probes used in the SMR analyses was limited. Moreover, the FDR approach used to correct for multiple testing could result in overlooking significant genes. Consequently, we could not rule out the possibility of missing some important genes.

Conclusions

In conclusion, we confirmed the involvement of *EVI2A* in the pathogenesis of EC and also identified additional genes that may contribute to the etiology of EC. Further research is essential to investigate the functions of these genes and to clarify the specific mechanisms underlying the etiology of EC.

Acknowledgments

Funding: This work was supported by National Natural Science Foundation of China (No. 82272724), Beijing Municipal Natural Science Foundation (No. Z220013); National High level Hospital Clinical Research Funding (No. 2022-PUMCH-A-067); National Institutes of Health/National Institute on Aging [Nos. P30AG10161, R01AG15819, R01AG17917, R01AG033678, R01AG36042, U01AG61356, and 1RF1AG064312–01 (to J.Y.)].

Footnote

Reporting Checklist: The authors have completed the STROBE-MR reporting checklist. Available at https://tcr.amegroups.com/article/view/10.21037/tcr-24-887/rc

Peer Review File: Available at https://tcr.amegroups.com/article/view/10.21037/tcr-24-887/prf

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at https://tcr.amegroups.com/article/view/10.21037/tcr-24-887/coif). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013).

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: https://creativecommons.org/licenses/by-nc-nd/4.0/.

References

1. Sung H, Ferlay J, Siegel RL, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin 2021;71:209-49.
2. Shisler R, Sinnott JA, Wang V, et al. Life after endometrial cancer: A systematic review of patient-reported outcomes. Gynecol Oncol 2018;148:403-13.
3. Cronin KA, Scott S, Firth AU, et al. Annual report to the nation on the status of cancer, part 1: National cancer statistics. Cancer 2022;128:4251-84.
4. Crosbie EJ, Kitson SJ, McAlpine JN, et al. Endometrial cancer. Lancet 2022;399:1412-28.
5. Ali AT. Risk factors for endometrial cancer. Ceska Gynekol 2013;78:448-59.
6. Win AK, Reece JC, Ryan S. Family history and risk of

endometrial cancer: a systematic review and meta-analysis. Obstet Gynecol 2015;125:89-98.
7. O'Mara TA, Glubb DM, Amant F, et al. Identification of nine new susceptibility loci for endometrial cancer. Nat Commun 2018;9:3166.
8. Painter JN, O'Mara TA, Marquart L, et al. Genetic Risk Score Mendelian Randomization Shows that Obesity Measured as Body Mass Index, but not Waist:Hip Ratio, Is Causal for Endometrial Cancer. Cancer Epidemiol Biomarkers Prev 2016;25:1503-10.
9. O'Mara TA, Glubb DM, Painter JN, et al. Comprehensive genetic assessment of the ESR1 locus identifies a risk region for endometrial cancer. Endocr Relat Cancer 2015;22:851-61.
10. Carvajal-Carmona LG, O'Mara TA, Painter JN, et al. Candidate locus analysis of the TERT-CLPTM1L cancer risk region on chromosome 5p15 identifies multiple independent variants associated with endometrial cancer risk. Hum Genet 2015;134:231-45.
11. Thompson DJ, O'Mara TA, Glubb DM, et al. CYP19A1 fine-mapping and Mendelian randomization: estradiol is causal for endometrial cancer. Endocr Relat Cancer 2016;23:77-91.
12. Painter JN, O'Mara TA, Batra J, et al. Fine-mapping of the HNF1B multicancer locus identifies candidate variants that mediate endometrial cancer risk. Hum Mol Genet 2015;24:1478-92.
13. Wang X, Glubb DM, O'Mara TA. 10 Years of GWAS discovery in endometrial cancer: Aetiology, function and translation. EBioMedicine 2022;77:103895.
14. Cheng TH, Thompson DJ, O'Mara TA, et al. Five endometrial cancer risk loci identified through genome-wide association analysis. Nat Genet 2016;48:667-74.
15. Lacey JV Jr, Yang H, Gaudet MM, et al. Endometrial cancer and genetic variation in PTEN, PIK3CA, AKT1, MLH1, and MSH2 within a population-based case-control study. Gynecol Oncol 2011;120:167-73.
16. Cheung LW, Hennessy BT, Li J, et al. High frequency of PIK3R1 and PIK3R2 mutations in endometrial cancer elucidates a novel mechanism for regulation of PTEN protein stability. Cancer Discov 2011;1:170-85.
17. Guan B, Mao TL, Panuganti PK, et al. Mutation and loss of expression of ARID1A in uterine low-grade endometrioid carcinoma. Am J Surg Pathol 2011;35:625-32.
18. Lax SF, Kendall B, Tashiro H, et al. The frequency of p53, K-ras mutations, and microsatellite instability differs in uterine endometrioid and serous carcinoma: evidence of distinct molecular genetic pathways. Cancer 2000;88:814-24.
19. Gasparri ML, Bellaminutti S, Farooqi AA, et al. Endometrial Cancer and BRCA Mutations: A Systematic Review. J Clin Med 2022;11:3114.
20. Meyer LA, Broaddus RR, Lu KH. Endometrial cancer and Lynch syndrome: clinical and pathologic considerations. Cancer Control 2009;16:14-22.
21. Risinger JI, Allard J, Chandran U, et al. Gene expression analysis of early stage endometrial cancers reveals unique transcripts associated with grade and histology but not depth of invasion. Front Oncol 2013;3:139.
22. Shi S, Tan Q, Feng F, et al. Identification of core genes in the progression of endometrial cancer and cancer cell-derived exosomes by an integrative analysis. Sci Rep 2020;10:9862.
23. Ren X, Liang J, Zhang Y, et al. Single-cell transcriptomic analysis highlights origin and pathological process of human endometrioid endometrial carcinoma. Nat Commun 2022;13:6300.
24. Tang X, Wang J, Tao H, et al. Regulatory patterns analysis of transcription factor binding site clustered regions and identification of key genes in endometrial cancer. Comput Struct Biotechnol J 2022;20:812-23.
25. Zhang YD, Hurson AN, Zhang H, et al. Assessment of polygenic architecture and risk prediction based on common variants across fourteen cancers. Nat Commun 2020;11:3353.
26. Mountjoy E, Schmidt EM, Carmona M, et al. An open approach to systematically prioritize causal variants and genes at all published human GWAS trait-associated loci. Nat Genet 2021;53:1527-33.
27. Zhu Z, Zhang F, Hu H, et al. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. Nat Genet 2016;48:481-7.
28. Feng H, Mancuso N, Gusev A, et al. Leveraging expression from multiple tissues using sparse canonical correlation analysis and aggregate tests improves the power of transcriptome-wide association studies. PLoS Genet 2021;17:e1008973.
29. Lloyd-Jones LR, Holloway A, McRae A, et al. The Genetic Architecture of Gene Expression in Peripheral Blood. Am J Hum Genet 2017;100:228-37.
30. Ghoussaini M, Mountjoy E, Carmona M, et al. Open Targets Genetics: systematic identification of trait-associated genes using large-scale genetics and functional genomics. Nucleic Acids Res 2021;49:D1311-20.
31. Yang Z, Yang J, Liu D, et al. Mendelian randomization analysis identified genes pleiotropically associated with

central corneal thickness. BMC Genomics 2021;22:517.
32. Gusev A, Ko A, Shi H, et al. Integrative approaches for large-scale transcriptome-wide association studies. Nat Genet 2016;48:245-52.
33. Liu Y, Chen S, Li Z, et al. ACAT: A Fast and Powerful p Value Combination Method for Rare-Variant Analysis in Sequencing Studies. Am J Hum Genet 2019;104:410-21.
34. GTEx Consortium; Laboratory, Data Analysis &Coordinating Center (LDACC)—Analysis Working Group; Statistical Methods groups—Analysis Working Group, et al. Genetic effects on gene expression across human tissues. Nature 2017;550:204-13.
35. Kho PF, Wang X, Cuéllar-Partida G, et al. Multi-tissue transcriptome-wide association study identifies eight candidate genes and tissue-specific gene expression underlying endometrial cancer susceptibility. Commun Biol 2021;4:1211.
36. Cawthon RM, O'Connell P, Buchberg AM, et al. Identification and characterization of transcripts from the neurofibromatosis 1 region: the sequence and genomic structure of EVI2 and mapping of other transcripts. Genomics 1990;7:555-65.
37. Cawthon RM, Andersen LB, Buchberg AM, et al. cDNA sequence and genomic structure of EV12B, a gene lying within an intron of the neurofibromatosis type 1 gene. Genomics 1991;9:446-60.
38. Qiu Z, Sun W, Gao S, et al. A 16-gene signature predicting prognosis of patients with oral tongue squamous cell carcinoma. PeerJ 2017;5:e4062.
39. Li S, Yang F, Yang YK, et al. Increased expression of ecotropic viral integration site 2A indicates a poor prognosis and promotes osteosarcoma evolution through activating MEK/ERK pathway. J Recept Signal Transduct Res 2019;39:368-72.
40. Chen HJ, Lin CM, Lee CY, et al. Kaempferol suppresses cell metastasis via inhibition of the ERK-p38-JNK and AP-1 signaling pathways in U-2 OS human osteosarcoma cells. Oncol Rep 2013;30:925-32.
41. Cheng S, Zhang X, Huang N, et al. Down-regulation of S100A9 inhibits osteosarcoma cell growth through inactivating MAPK and NF-κB signaling pathways. BMC Cancer 2016;16:253.
42. Wu R, Li D, Tang Q, et al. A Novel Peptide from Vespa ducalis Induces Apoptosis in Osteosarcoma Cells by Activating the p38 MAPK and JNK Signaling Pathways. Biol Pharm Bull 2018;41:458-64.
43. Buchberg AM, Bedigian HG, Jenkins NA, et al. Evi-2, a common integration site involved in murine myeloid leukemogenesis. Mol Cell Biol 1990;10:4658-66.
44. Li XW, Rees JS, Xue P, et al. New insights into the DT40 B cell receptor cluster using a proteomic proximity labeling assay. J Biol Chem 2014;289:14434-47.
45. Philpott C, Tovell H, Frayling IM, et al. The NF1 somatic mutational landscape in sporadic human cancers. Hum Genomics 2017;11:13.
46. Ruth KS, Day FR, Tyrrell J, et al. Using human genetics to understand the disease impacts of testosterone in men and women. Nat Med 2020;26:252-8.
47. Xu J, Xu T, Wu B, et al. Structure of sorting nexin 11 (SNX11) reveals a novel extended phox homology (PX) domain critical for inhibition of SNX10-induced vacuolation. J Biol Chem 2013;288:16598-605.
48. Qin B, He M, Chen X, et al. Sorting nexin 10 induces giant vacuoles in mammalian cells. J Biol Chem 2006;281:36891-6.
49. Chandra M, Chin YK, Mas C, et al. Classification of the human phox homology (PX) domains based on their phosphoinositide binding specificities. Nat Commun 2019;10:1528.
50. O'Mara TA, Spurdle AB, Glubb DM, et al. Analysis of Promoter-Associated Chromatin Interactions Reveals Biologically Relevant Candidate Target Genes at Endometrial Cancer Risk Loci. Cancers (Basel) 2019;11:1440.
51. Minczuk M, He J, Duch AM, et al. TEFM (c17orf42) is necessary for transcription of human mtDNA. Nucleic Acids Res 2011;39:4284-99.
52. Chakrabarty RP, Chandel NS. Beyond ATP, new roles of mitochondria. Biochem (Lond) 2022;44:2-8.
53. Wan L, Wang Y, Zhang Z, et al. Elevated TEFM expression promotes growth and metastasis through activation of ROS/ERK signaling in hepatocellular carcinoma. Cell Death Dis 2021;12:325.

Cite this article as: Zhang G, Mao S, Yuan G, Wang Y, Yang J, Dai Y. Exploring potential causal genetic variants and genes for endometrial cancer: Open Targets Genetics, Mendelian randomization, and multi-tissue transcriptome-wide association analysis. Transl Cancer Res 2024;13(11):5971-5982. doi: 10.21037/tcr-24-887

© AME Publishing Company.