Peer Reviewer File

Article information: https://dx.doi.org/10.21037/tcr-24-887

Reviewer A

The authors have used GWAS summary statistics with a variety of approaches to identify candidate causal genes underlying endometrial cancer (EC) risk. The study integrates Open Targets Genetics (OTG), Summary-based Mendelian Randomization (SMR), and multi-tissue transcriptome-wide association study (TWAS) methodologies. Two genes, *EVI2A* and *SNX11*, were identified by all three approaches.

Response: Thank you for taking the time to review our manuscript. We sincerely appreciate your insightful feedback, which has been instrumental in improving the quality of our work. In response to your valuable suggestions, we have made the necessary revisions to the manuscript.

Major Issues

1. Limited Interpretation and Biological Relevance: The authors have highlighted several genes, but there has been limited interpretation of their effect on EC. There is no discussion on the differences between the approaches used and the significance of the gene selection process. There is no direction of effect provided for gene expression on EC risk, which diminishes the potential impact of the findings on advancing the understanding of EC biology. Furthermore, EVI2A and SNX11 have already been established as candidate endometrial cancer susceptibility genes in previous studies (e.g., O'Mara et al. 2019, Kho et al. 2021). The current study confirms these findings but does not significantly advance the field. This should be clearly stated to provide an accurate context for the study's contribution. The suggestion that the authors have identified genes crucial to EC pathogenesis is premature without further functional validation.

Response: Thank you for the thoughtful comments. Knowledge about the effects of some of the genes identified in our study is limited. That said, we expanded the discussion section to provide a deeper discussion of the effects of these genes on EC. In the discussion section, we added a paragraph highlighting the differences between OTG, SMR, and sCCA+ACAT (page 12). Due to the cross-tissue nature of the multi-tissue TWAS, it provided only tissue-specific test statistics (**Table S3**) without overall effect size and direction. We have revised the manuscript to clarify this point (see our response to comment 4).

Our study indeed confirms the involvement of *EVI2A* in the pathogenesis of EC through various bioinformatical approaches. We also aim to explore additional potential genes that may influence EC susceptibility. In the revised manuscript, we reiterate that *EVI2A* merely confirms previous findings by O'Mara et al. and Kho et al. (page 12). Moreover, we emphasize our intention to explore novel genes associated with EC to deepen our understanding of its genetic basis (page 6).

We agree that it is premature to suggest the identification of genes crucial to EC pathogenesis without functional validation. Accordingly, we have moderated our language throughout the manuscript, including in the Abstract and the main text. We now emphasize that while our findings point to potential candidate genes, further functional studies are essential to validate their role in endometrial cancer pathogenesis (page 15). This adjustment ensures that our conclusions remain appropriately cautious and aligned with the evidence.

2. Contextual Background: The "What is known and what is new?" section lacks depth and does not adequately discuss the current state of the field. It fails to reference key reviews of endometrial cancer GWAS or highlight the contributions of previous post-GWAS studies. This deficiency reflects a lack of contextual grounding and diminishes the perceived novelty of the study. The introduction

should be expanded to include a thorough review of existing literature, emphasizing what is already known about the genetic mechanisms underlying EC and how this study builds upon those findings.

Response: Thank you for your insightful comments. We have completely revised the Highlight Box to concisely summarize the existing genetic findings on endometrial cancer (EC), the contributions of our study, and the implications of our findings. Please note that due to the journal's 250-word limit for this section, we have focused on a brief summary of the literature (page 3). Additionally, we significantly expanded the introduction section to provide a more comprehensive review of existing research (pages 5-6).

3. Reference to EC GWAS: The EC GWAS that forms the basis for all three approaches has not been clearly referenced throughout. It should be explicitly stated that the GWAS data from O'Mara et al. 2018 underpins the analysis presented in the manuscript.

   Response: The reviewer correctly noted that the EC GWAS data from O'Mara et al. 2018 underpins all three analytical approaches in our study. In the discussion section, we emphasized this foundation before elaborating on the differences between the three approaches (page 12).

4. Discussion of Methodological Limitations: While the sCCA TWAS+ACAT method could have increased power, there should be some discussion of how multi-tissue TWAS approaches can impede the detection of genetic mechanisms that affect EC risk by acting in specific organs and tissues. TWAS genetic associations that are not tissue-specific may be more difficult to pursue for prevention and treatment. The direction of effect for genes identified by TWAS analysis should be provided as it is a clear limitation of the work performed.

Response: Thank you for your thoughtful feedback regarding our use of sCCA TWAS+ACAT. This method was employed to complement the SMR analysis. We opted for sCCA TWAS+ACAT over tissue-specific TWAS (i.e., uterus) due to the limited sample size of available uterine eQTL data, which would have resulted in insufficient statistical power (please also see our response to comment 6). In contrast, sCCA TWAS+ACAT offers substantially higher power than traditional single-tissue TWAS methods for identifying genes with genetically predicted expression associated with traits. However, we acknowledge that sCCA TWAS+ACAT may not fully capture tissue-specific genetic mechanisms in EC pathogenesis. We have included this limitation in the discussion section (page 15).

We agree that providing the direction of effect for genes identified by TWAS is valuable. Due to its cross-tissue nature, the multi-tissue TWAS provided only tissue-specific test statistics (**Table S3**) without overall effect sizes or directions. We have emphasized this limitation in the manuscript (page 15).

5. SMR Analysis Concerns: The SMR results show likely heterogeneity for SKAP1 and SNX11, as evidenced by low HEIDI p-values, suggesting likely LD-contamination and rendering these results invalid.

Response: Thank you for highlighting this important point. We agree that the low HEIDI p-values likely indicate linkage, meaning the pleiotropic association could result from the top associated cis-eQTL being in linkage disequilibrium (LD) with two distinct causal variants—one affecting gene expression and the other trait variation. Such findings may be less functionally relevant. Based on your comments, we revised the manuscript, including updates to the abstract, Results section, and Discussion section. Specifically, we addressed the HEIDI test results and their implications (page 10) and revised our major findings throughout the manuscript.

6. Clarity on eQTL Data: The CAGE eQTL data is not clearly defined – it is unclear which tissue is being interrogated. The provided website points to a page with many datasets, making it difficult to determine which one has been used in this study specifically.

   Response: We apologize for the lack of clarity regarding the specific tissue source of the CAGE eQTL data. The CAGE eQTL data are derived from peripheral blood. While we acknowledge that eQTL data are tissue-specific, we chose CAGE eQTL data due to its relatively large sample size compared to tissue-specific eQTL data. For instance, the uterus eQTL data in GTEx V8 has a sample size of fewer than 150. We have added this information in the Methods section (page 7) and as a limitation in the Discussion section (page 15). Additionally, we provided a direct download link for the eQTL data (page7).

7. Data Reproducibility: More information should be provided about LD pruning and allele frequency checking ("Basic information of summarized data"). The work cannot be recapitulated as it stands.

   Response: We fully agree that data reproducibility is crucial in SMR studies. To address this, we added a supplementary table detailing all the settings used in the SMR analysis, including LD pruning and allele frequency checks (page 8, **Table S1**). Note that we used default settings for both SMR and cross-tissue TWAS analyses. We believe that our results can be replicated relatively easily if researchers use the same GWAS and eQTL data as outlined in our manuscript.

8. Statistical Corrections: The use of FDR is not a limitation – Bonferroni correction is recommended. The authors are being liberal using an FDR.

   Response: Thank you for your insightful comment regarding multiple testing correction. We understand the importance of stringent correction methods for validating our findings. We employed the False Discovery Rate (FDR) correction

method to balance the discovery of true positives while controlling the rate of false positives. Given the exploratory nature of our study, the FDR approach is well-suited for identifying relevant associations without being overly conservative. While the Bonferroni correction is more stringent and reduces the likelihood of Type I errors, it can be overly conservative in SMR analyses, potentially increasing the risk of Type II errors and missing true pleiotropic associations. To address your concern, we re-evaluated our significant findings using the Bonferroni correction. Two genes (*EVI2A* and *SKAP1*) remained significant after Bonferroni correction, with *EVI2A* passing the HEIDI test. Thus, *EVI2A* is the gene that survived the stringent Bonferroni correction and passed the HEIDI test. Additionally, Bonferroni correction in the TWAS+ACAT analysis confirmed the same two genes (*EVI2A* and *SKAP1*) as significantly associated with EC. Considering your other comments regarding the HEIDI test, this indicates that our primary finding remained consistent with Bonferroni correction (i.e., only *EVI2A* showed significant pleiotropic association with EC). We added this finding to the Results section (page 10).

Minor Issues

Figures and Tables:

•Table 1: Needs revision for clarity. The reason for the difference in the number of genetic variants available for the EC GWAS dataset should be explained.

Response: Thank you for your careful comments. While both SMR and cross-tissue TWAS utilized the same EC GWAS data from O'Mara, the number of eligible genetic variants differs due to the distinct filtering procedures applied. Specifically, the number of genetic variants for SMR reflects the SNPs that passed the initial filtering criteria outlined in **Table S1**. In contrast, the number of variants in cross-tissue TWAS represents the total potentially eligible genetic variants analyzed across different tissues. We have added a note to Table 1 to clarify this distinction.

•Figure 2: No solid rhombuses in the plot – not surprising given the significant HEIDI p-value. These genes shouldn't be reported as associated with EC risk based on SMR analysis.

Response: We agree with the reviewer's suggestion. As mentioned earlier, we have revised the manuscript regarding the pleiotropic associations of genes that did not pass the HEIDI test. We highlighted that *EVI2A* was the only gene identified by all three methods. We have replaced the original Figure 2 with Figure S1 and updated the legends in the supplementary figures to indicate that hollow rhombuses represent probes that did not pass the HEIDI test.

Textual Errors:

•There are several errors in the text (e.g., SVI2A instead of EVI2A under "Pleiotropic association with EC").

Response: We apologize for the typos in the original submission. We have thoroughly reviewed the entire manuscript and made the necessary corrections.

Reviewer B

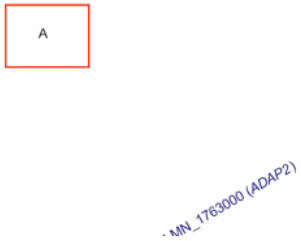
1. Figures and tables

- Figure 1 legend: Please check if it should be OTG.

  OGT, Open Targets Genetics;

  Response: You are correct; the abbreviation should indeed be OTG. This change has been made.

- Please check if there is figure 2B in Figure 2. If not, please remove A in the Figure.



Response: The letter 'a' should not have appeared in the figure. This was generated automatically by our program. We have removed it and checked and corrected the supplementary figures accordingly.

- Table 2: It should be “OTG”.

OGT, Open Genetics Target; LD, linkage disequilibrium

Response: Yes, it should be OTG. We made the change.

- Table 3: There is no [a] in the table but its explanation in the foot. Please check and revise.

[a]P_{eQTL} is the P-value
analysis. Beta is the

Response: We removed a from the table note.

- There are two Table S1 in the supplementary file you submitted. Please check and revise.

Table S1. Genes showing potential pleiotropic association with EC using CAGE eQTL data

probeID	ProbeChr	Gene	Probe_bp	topSNP	topSNP_ch	topSNP_bp	A1	A2	Freq	b_GWAS	se_GWAS	p_GWAS
ILMN_1751400	17	SKAP1	46210873	rs2938483	17	46318004	C	T	0.347561	-0.0886	0.015646	1.49E-08
ILMN_2369018	17	EVI2A	29644832	rs7505	17	29644852	G	A	0.359756	0.074118	0.015541	1.85E-06
ILMN_1809347	15	SRP14	40328444	rs1772252	15	40320267	G	A	0.378049	0.08769	0.015726	2.46E-08
ILMN_1733579	17	EVI2A	29645189	rs2525570	17	29681245	A	G	0.371951	0.07104	0.015533	4.80E-06
ILMN_1683950	17	SNX11	46200008	rs6206495	17	46193967	C	T	0.189024	-0.08302	0.018841	1.05E-05
ILMN_1696051	17	SNX11	46188096	rs1294987	17	46158050	C	T	0.189024	-0.08257	0.018836	1.17E-05

Supplementary Table 1. Default parameters in the SMR analyses

parameter	meaning
--maf	removes SNPs based on a minor allele frequency (MAF) threshold in the reference sample
--diff-freq	excludes SNPs with allele frequency differences between any pairwise data sets larger than the specified threshold
--diff-freq-prop	stops the analysis if the proportion of SNPs being excluded by --diff-freq is larger than the specified value
--peqtl-smr	p-value threshold to select the top associated eQTL for the SMR test
--peqtl-heidi	threshold of eQTL p-value to select eQTLs for the HEIDI test
--ld-upper-limit	upper LD r-squared threshold used to prune SNPs (eQTLs) in the HEIDI test

- The titles of the following table are inconsistent. Please check and revise.

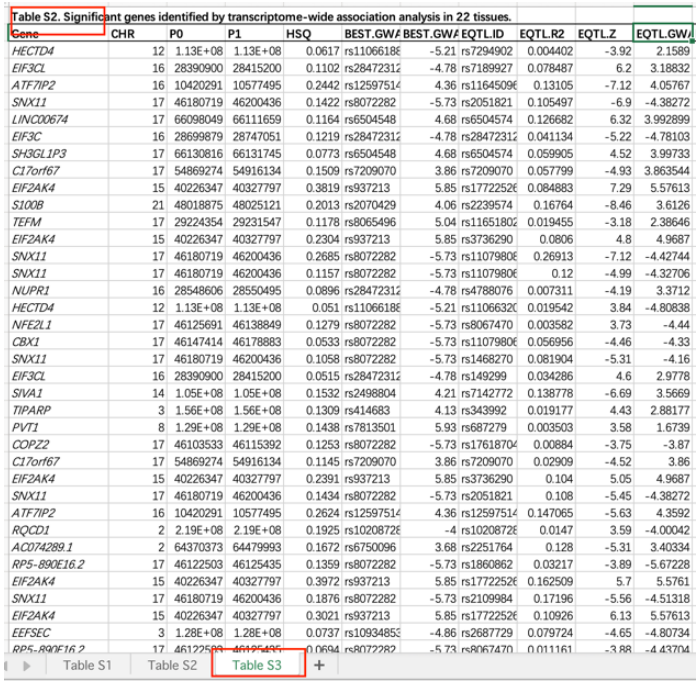


Table S2. Significant genes identified by transcriptome-wide association analysis in 22 tissues.

Gene	CHR	P0	P1	HSQ	BEST.GWA	BEST.GWA	EQTL.ID	EQTL.R2	EQTL.Z	EQTL.GWA
HECTD4	12	1.13E+08	1.13E+08	0.0617	rs11066188	-5.21	rs7294902	0.004402	-3.92	2.1589
EIF3CL	16	28390900	28415200	0.1102	rs28472312	-4.78	rs7189927	0.078487	6.2	3.18832
ATF7IP2	16	10420291	10577495	0.2442	rs12597514	4.36	rs11645096	0.13105	-7.12	4.05767
SNX11	17	46180719	46200436	0.1422	rs8072282	-5.73	rs2051821	0.105497	-6.9	-4.38272
LINC00674	17	66098049	66111659	0.1164	rs6504548	4.68	rs6504574	0.126682	6.32	3.992899
EIF3C	16	28699879	28747051	0.1219	rs28472312	-4.78	rs28472312	0.041134	-5.22	-4.78103
SH3GL1P3	17	66130816	66131745	0.0773	rs6504548	4.68	rs6504574	0.059905	4.52	3.99733
C17orf67	17	54869274	54916134	0.1509	rs7209070	3.86	rs7209070	0.057799	-4.93	3.863544
EIF2AK4	15	40226347	40327797	0.3819	rs937213	5.85	rs17722526	0.084883	7.29	5.57613
S100B	21	48018875	48025121	0.2013	rs2070429	4.06	rs2239574	0.16764	-8.46	3.6126
TEFM	17	29224354	29231547	0.1178	rs8065496	5.04	rs11651802	0.019455	-3.18	2.38646
EIF2AK4	15	40226347	40327797	0.2304	rs937213	5.85	rs3736290	0.0806	4.8	4.9687
SNX11	17	46180719	46200436	0.2685	rs8072282	-5.73	rs11079808	0.26913	-7.12	-4.42744
SNX11	17	46180719	46200436	0.1157	rs8072282	-5.73	rs11079806	0.12	-4.99	-4.32706
NUPR1	16	28548606	28550495	0.0896	rs28472312	-4.78	rs4788076	0.007311	-4.19	3.3712
HECTD4	12	1.13E+08	1.13E+08	0.051	rs11066188	-5.21	rs11066320	0.019542	3.84	-4.80838
NFE2L1	17	46125691	46138849	0.1279	rs8072282	-5.73	rs8067470	0.003582	3.73	-4.44
CBX1	17	46147414	46178883	0.0533	rs8072282	-5.73	rs11079806	0.056956	-4.46	-4.33
SNX11	17	46180719	46200436	0.1058	rs8072282	-5.73	rs1468270	0.081904	-5.31	-4.16
EIF3CL	16	28390900	28415200	0.0515	rs28472312	-4.78	rs149299	0.034286	4.6	2.9778
SIVA1	14	1.05E+08	1.05E+08	0.1532	rs2498804	4.21	rs7142772	0.138778	-6.69	3.5669
TIPARP	3	1.56E+08	1.56E+08	0.1309	rs414683	4.13	rs343992	0.019177	4.43	2.88177
PVT1	8	1.29E+08	1.29E+08	0.1438	rs7813501	5.93	rs687279	0.003503	3.58	1.6739
COPZ2	17	46103533	46115392	0.1253	rs8072282	-5.73	rs17618704	0.00884	-3.75	-3.87
C17orf67	17	54869274	54916134	0.1145	rs7209070	3.86	rs7209070	0.02909	-4.52	3.86
EIF2AK4	15	40226347	40327797	0.2391	rs937213	5.85	rs3736290	0.104	5.05	4.9687
SNX11	17	46180719	46200436	0.1434	rs8072282	-5.73	rs2051821	0.108	-5.45	-4.38272
ATF7IP2	16	10420291	10577495	0.2624	rs12597514	4.36	rs12597514	0.147065	-5.63	4.3592
RQCD1	2	2.19E+08	2.19E+08	0.1925	rs10208728	-4	rs10208728	0.0147	3.59	-4.00042
AC074289.1	2	64370373	64479993	0.1672	rs6750096	3.68	rs2251764	0.128	-5.31	3.40334
RP5-890E16.2	17	46122503	46125435	0.1359	rs8072282	-5.73	rs1860862	0.03217	-3.89	-5.67228
EIF2AK4	15	40226347	40327797	0.3972	rs937213	5.85	rs17722526	0.162509	5.7	5.5761
SNX11	17	46180719	46200436	0.1876	rs8072282	-5.73	rs2109984	0.17196	-5.56	-4.51318
EIF2AK4	15	40226347	40327797	0.3021	rs937213	5.85	rs17722526	0.10926	6.13	5.57613
EEFSEC	3	1.28E+08	1.28E+08	0.0737	rs10934853	-4.86	rs2687729	0.079724	-4.65	-4.80734
RP5-890E16.2	17	46122503	46125435	0.0694	rs8072282	-5.73	rs8067470	0.011161	-3.88	-4.43704

Table S1 | Table S2 | Table S3 | +

Response: We apologize for the oversight. In the previous revision, we added a supplementary table but neglected to update the title. The order has now been corrected, and the format has been made consistent throughout.

2. Please check if any references should be cited since you mentioned **studies**.
   - OTG also reports other traits that colocalize with EC at a given locus based on previous GWAS **studies**.

   Response: A citation is not required here, as the statement simply describes the function of OTG.

   - Further, *SNX11* has been identified as a potential target for EC risk variation through enhancer-promoter chromatin looping **studies**(50).

   Response: To enhance clarity, we have replaced the term “studies” with “analysis” (page 14).