


Phenome-wide association study of ovarian cancer identifies common comorbidities and reveals shared genetics with complex diseases and biomarkers

Anwar Mulugeta^{1,2,3}  | Amanda L. Lumsden^{1,2} | Iqbal Madakkattel^{1,2} | David Stacey^{1,2} | S. Hong Lee^{1,2,4} | Johanna Mäenpää^{5,6} | Martin K. Oehler^{7,8} | Elina Hyppönen^{1,2}

¹Australian Centre for Precision Health, Unit of Clinical and Health Sciences, University of South Australia, Adelaide, South Australia, Australia

²South Australian Health and Medical Research Institute, Adelaide, South Australia, Australia

³Department of Pharmacology and Clinical Pharmacy, College of Health Science, Addis Ababa University, Addis Ababa, Ethiopia

⁴UniSA Allied Health & Human Performance, University of South Australia, Adelaide, South Australia, Australia

⁵Faculty of Medicine and Medical Technology, Tampere University, Tampere, Finland

⁶Cancer Centre, Tampere University and University Hospital, Tampere, Finland

⁷Department of Gynaecological Oncology, Royal Adelaide Hospital, Adelaide, South Australia, Australia

⁸Adelaide Medical School, Robinson Research Institute, University of Adelaide, Adelaide, South Australia, Australia

Correspondence

Anwar Mulugeta, Australian Centre for Precision Health, c/o SAHMRI, University of South Australia, GPO Box 2471, Adelaide SA 5001, Australia.
Email: anwarmulugeta.gebremichael@unisa.edu.au

Funding information

Medical Research Future Fund of Australia, Grant/Award Number: MRF2007431

Abstract

Background: Ovarian cancer (OC) is commonly diagnosed among older women who have comorbidities. This hypothesis-free phenome-wide association study (PheWAS) aimed to identify comorbidities associated with OC, as well as traits that share a genetic architecture with OC.

Methods: We used data from 181,203 white British female UK Biobank participants and analysed OC and OC subtype-specific genetic risk scores (OC-GRS) for an association with 889 diseases and 43 other traits. We conducted PheWAS and colocalization analyses for individual variants to identify evidence for shared genetic architecture.

Results: The OC-GRS was associated with 10 diseases, and the clear cell OC-GRS was associated with five diseases at the FDR threshold ($p = 5.6 \times 10^{-4}$). Mendelian randomization analysis (MR) provided robust evidence for the association of OC with higher risk of “secondary malignant neoplasm of digestive systems” (OR 1.64, 95% CI 1.33, 2.02), “ascites” (1.48, 95% CI 1.17, 1.86), “chronic airway obstruction” (1.17, 95% CI 1.07, 1.29), and “abnormal findings on examination of the lung” (1.51, 95% CI 1.22, 1.87). Analyses of lung spirometry measures provided further support for compromised respiratory function. PheWAS on individual OC variants identified five genetic variants associated with other diseases,

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2024 The Authors. *Cancer Medicine* published by John Wiley & Sons Ltd.

and seven variants associated with biomarkers (all, $p \leq 4.5 \times 10^{-8}$). Colocalization analysis identified rs4449583 (from *TERT* locus) as the shared causal variant for OC and seborrheic keratosis.

Conclusions: OC is associated with digestive and respiratory comorbidities. Several variants affecting OC risk were associated with other diseases and biomarkers, with this study identifying a novel genetic locus shared between OC and skin conditions.

KEYWORDS

comorbidities, ovarian cancer, PheWAS, *TERT*, UK biobank

1 | INTRODUCTION

Ovarian cancer (OC) is a notorious gynecological disorder, with only 44% of affected women surviving 5 years.¹ Reducing mortality associated with OC remains a challenge as the poor understanding of the origin and biology of the disease limits opportunities to prevention, and the disease is typically detected at an advanced stage when treatment options are limited.² Approximately 90% of OC are epithelial OC and arise from epithelial cells, predominantly of the fallopian tube and endometrium.²⁻⁴ Epithelial OC can be further classified into serous, endometrioid, clear cell, and mucinous subtypes, based on their histological characteristics³ and indeed, the pathogenesis, disease progression, and risk factors are known to vary between OC subtypes.^{3,5} Given that OC is commonly diagnosed in women aged 50 years and above,⁶ many patients present with at least one comorbidity, which can influence treatment decisions, impact prognosis and affect overall survival.^{7,8} Hence, exploring the associations between OC and comorbidities, along with understanding potential shared genetic links between OC and other traits, may offer insights into the underlying etiology of the disease. Such insights could lead to facilitating earlier detection and better management of the disease.

Phenome-wide association study (PheWAS) is a powerful approach for testing the association of genetic variants or their risk score with a broad range of phenotypes and disease outcomes.⁹ This approach offers the advantage of being hypothesis-free, allowing it to explore numerous associations without constraints of the number of outcomes. Consequently, PheWAS has the potential to unveil novel genetic associations with diseases.⁹ In this study we use data on genetic OC risk variants and 889 disease outcomes from 181,203 female UK Biobank participants to identify comorbidities and clinical characteristics associated with the genetic risk of OC. We test for evidence of a direct association between OC and possible comorbidities using Mendelian randomization (MR), an approach less susceptible to confounding and reverse causation,¹⁰ and extend these analyses

to focus on the genetic risk reflecting distinct histological OC subtypes. We also use the PheWAS approach to explore each OC associated variant individually, which can provide evidence for shared genetic contributions, and provide insight into the underlying disease mechanisms.

2 | METHODS

2.1 | Study population

UK Biobank is a prospective cohort study of half a million participants with rich phenotype and genotype data collected across 22 centres in England, Scotland, and Wales. The baseline assessment was conducted between 2006 and 2010 when the participants were aged between 37 and 73 years.¹¹ A wide range of data was collected, including sociodemographic information, lifestyle factors, physical measures, blood and urine samples for biomarker, and genetic profiling, with continuing updates through follow-up assessments and linkage with electronic health records and death registrations.¹¹ For the primary analyses in this study, we restricted the sample to 181,203 female participants who were of white British ancestry and were not genetically related (Figure S1) and included male participants in sensitivity analyses ($n = 156,220$). The study was conducted under UK Biobank application number 89630. UK Biobank obtained informed consent from all participants to collect and use linked data for future research use, and the UK Biobank study has ethical approval from the National Information Governance Board for Health and Social Care and Northwest Multicenter Research Ethics Committee (11/NW/0382).

2.2 | Disease outcomes and biomarkers

We used International Classification of Diseases (ICD version 9 and 10) codes from hospital episode statistics

and death registrations to identify disease outcomes (Supplementary Methods: Appendix S1). We mapped ICD-9/10 codes to phecodes, which are composite groupings reflecting clinically relevant disease outcomes.¹² We defined individuals with the pcode-of-interest as cases, and those without the pcode-of-interest, or any other phecodes from the same disease category as controls (Supplementary Methods: Appendix S1).^{12,13} We included all phecodes with at least 200 cases in our analyses,¹⁴ resulting in 889 phecodes from 18 disease categories (Table S1).

We also included 30 serum and four urine biomarkers from the baseline assessment, which are commonly used for the diagnosis and monitoring of chronic diseases, as outcome measures (Supplementary Methods: Appendix S1). Additional analyses were conducted against selected baseline physiological characteristics (namely body mass index [BMI], body fat percentage [BFP], basal metabolic rate [BMR], blood pressures, and measures of lung function [spirometry]) due to their known associations with comorbidity risks. The spirometry measures included forced expiratory volume in 1-second (FEV1), forced vital capacity (FVC), FEV1/FVC ratio, and peak expiratory flow (PEF). We applied log-transformation to biomarkers and physiological measures to mitigate the skewness of the distribution, and further scaled to a mean of zero and a standard deviation (SD) of one to allow us to present the results using the same scale for all measures.

2.3 | Genetic variants and genetic risk score

We selected 35 genetic variants associated with OC from previous genome-wide association studies (GWAS), including the 12 novel variants identified in the most recent GWAS¹⁵ and 23 variants identified in earlier GWASs (minor allele frequency $\geq 1.4\%$, Table S2, Figure S2).^{16–23} We constructed the genetic risk score (GRS) using 31 genetic variants after excluding one variant that had no proxy at $r^2 \geq 0.8$ (rs555025179), and three variants (rs12131772, rs4691139, and rs12938171) that were not replicated in the recent GWAS,¹⁵ which we extracted from the imputed genotype UK Biobank data²⁴ (Table S2). The GRS was constructed for each individual by summing the weighted number of OC risk-associated variants, with the weight taken from the recent OC GWAS.¹⁵ We also constructed risk scores for each OC subtype, incorporating 20, five, five, and eight variants for serous, endometrioid, clear cell, and mucinous OC subtypes, respectively (Supplementary Methods: Appendix S1).

2.4 | Statistical analyses

We conducted phenome-wide analyses using R-package PheWAS²⁵ which allowed us to fit logistic regression (889 disease outcomes) or linear regression (43 biomarkers/physiological characteristics) against the OC GRS with all models adjusted for age, assessment centre, type of genotyping array, birth location, and 40 principal components (Figure S2). We repeated the phenome-wide analyses for each OC subtype GRS to identify subtype-specific disease associations, and further, conducted the analyses in male participants to understand whether the OC—disease associations involve nonsex-specific mechanistic pathways.

The phenome-wide association signals that passed the FDR threshold ($p < 5.6 \times 10^{-4}$ for the diseases and $p < 0.02$ for the biomarkers and physiological measures, Supplementary Methods: Appendix S1) were taken forward to MR analysis to examine evidence for direct association with OC. We used two-sample MR approach where variant—OC association estimates were taken from the recent OC GWAS,¹⁵ and the variant-outcome association estimates were from the UK Biobank, where the exposure and outcome data sources had no sample overlap. MR analyses were conducted on signals associated with the overall OC GRS or with any subtype-based GRSs. We used inverse variance weighted (IVW) MR²⁶ as the primary method and conducted sensitivity analyses using weighted median,²⁷ weighted mode,²⁸ MR-PRESSO,²⁹ and MR-Egger,³⁰ with each method working on different pleiotropic assumptions (Supplementary Methods: Appendix S1). We tested for horizontal pleiotropy using MR-Egger intercept, MR-PRESSO outlier and distortion tests, and leave-one-out and leave-block-out methods (Table S3, Supplementary Methods: Appendix S1). MR analyses were repeated excluding the OC cases. To test the MR assumption that the genetic instrument should not be associated with potential confounding factors, we explored the association between the OC GRSs and age, education, Townsend deprivation index, BMI, physical activity, and alcohol consumption (Appendix S1, Table S4). For reporting the MR analysis, we followed the STROBE-MR guidelines.³¹

We repeated the phenome-wide analyses for each OC genetic variant individually, again conducting analyses separately for men. We assessed the novelty of the PheWAS findings by checking the reported associations in the GWAS Catalog. We considered a locus-trait association as potentially novel if neither the lead variant nor any variant in linkage disequilibrium (LD) with the lead variant (with an $r^2 \geq 0.1$) had previously been reported to be associated with the same or related traits. For those potentially novel associations, we conducted a follow-up colocalization analysis using the HyPrColoc (Hypothesis Prioritization in multi-trait Colocalization) tool to assess

whether there is evidence of shared causal genetic variants between pairs of traits.³² We considered a posterior probability of full colocalization (PPFC) >0.64 as indicative of shared genetic etiology between these traits (Supplementary Methods: Appendix S1).³²

Statistical analyses were performed using STATA SE version 17.1, R version 4.2.1 and PLINK2.³³

3 | RESULTS

We included up to 181,203 female white British participants in the primary analyses, of whom 1863 (1.3%) had been diagnosed with OC (Table 1). The prevalence of OC was higher among participants who were older, obese, never drinkers of alcohol, and those with poor general health status at the baseline (for all, prevalence $\geq 1.2\%$,

and $p < 0.03$). The OC risk GRS was normally distributed and associated with OC as expected, with stronger association in the highest compared to the lowest genetic risk group (F -statistics = 23.7, $p = 4.4 \times 10^{-16}$, Figure S3). No associations were found between the risk scores and potential confounders after correction for multiple testing (Table S5).

3.1 | Phenome-wide-based Mendelian randomization

We investigated 889 disease outcomes from 18 disease categories in the phenome-wide analyses. We observed significant associations between the OC GRS and 10 diseases from four disease categories (neoplasms, digestive, respiratory, and genitourinary diseases) after FDR correction

TABLE 1 Characteristics of participants by ovarian cancer status.

		All, n (%)	Ovarian cancer, n (%)	Controls, n (%)
Age (in years)	39–49	40,116 (22.2)	231 (12.4)	39,885 (22.2)
	50–59	62,041 (34.2)	556 (29.8)	61,485 (34.3)
	60–73	79,046 (43.6)	1076 (57.8)	77,970 (43.5)
<i>P</i>			1.1×10^{-51}	
Education	None	31,059 (17.1)	421 (22.6)	30,638 (17.1)
	NVQ/CSE/A-levels	83,890 (46.3)	819 (44.0)	83,071 (46.3)
	Degree/professional	64,785 (35.8)	607 (32.6)	64,178 (35.8)
	Missing	1469 (0.8)	16 (0.8)	1453 (0.8)
<i>P</i>			0.06	
BMI (kg/m ²)	<18.5	1324 (0.7)	10 (0.5)	1314 (0.7)
	18.5–<25	70,988 (39.2)	643 (34.5)	70,345 (39.2)
	25–<30	66,596 (36.8)	713 (38.3)	65,883 (36.8)
	≥ 30	41,746 (23.0)	489 (26.3)	41,257 (23.0)
	Missing	549 (0.3)	8 (0.4)	541 (0.3)
<i>P</i>			0.005	
Alcohol consumption	Never	7808 (4.3)	109 (5.9)	7699 (4.3)
	Previous	6526 (3.6)	64 (3.4)	6462 (3.6)
	Current	166,747 (92.0)	1690 (90.7)	165,057 (92.0)
	Missing	122 (0.1)	0 (0.0)	122 (0.1)
<i>P</i>			0.03	
General health	Excellent	31,630 (17.5)	286 (15.4)	31,344 (17.5)
	Good	109,012 (60.2)	1041 (55.9)	107,971 (60.2)
	Fair	33,519 (18.5)	405 (21.7)	33,114 (18.5)
	Poor	6431 (3.5)	112 (6.0)	6319 (3.5)
	Missing	611 (0.3)	19 (1.0)	592 (0.3)
<i>P</i>			1.0×10^{-16}	

Abbreviation: BMI, body mass index.

Note: *p* is *p* value generated from likelihood ratio test in logistic models adjusted for age, birth location, and assessment centers.

($p = 5.6 \times 10^{-4}$). The top signals were for “cancer of the female genital organs”, “malignant neoplasm of ovary and other uterine adnexa”, and “malignant neoplasm of the ovary”, which confirms the “relevance assumption” of the MR approach¹⁰ (Supplementary Methods: Appendix S1) for using the GRS as instrument for OC (for all, $p = 4.8 \times 10^{-4}$, Figure 1, Panel A). The GRS was also associated with higher risk of other cancers and abnormal growths including “secondary malignant neoplasm of digestive systems”, “lipoma, including lipoma of skin and subcutaneous tissue” ($p = 3.9 \times 10^{-4}$), and “ovarian cyst” ($p = 6.0 \times 10^{-5}$). In addition, we identified associations with respiratory outcomes including “chronic airway obstruction” and “abnormal findings on examination of the lungs”, and with the digestive system outcome, “ascites” ($p = 3.2 \times 10^{-4}$).

In analyses using GRSs for specific OC subtypes (Figure 1 Panel B and C), the clear cell OC GRS was associated with additional diseases and increased risks such as “benign neoplasm of uterus”, and “uterine leiomyoma”, “disorders of menstruation”, “gram negative septicemia”, and “postoperative infection” (for all, $p \leq 2.3 \times 10^{-4}$). Overall, in addition to overall OC-related outcomes, 12 other disease outcomes were associated with at least one form of the OC GRS. Notably, none of these associations were detected when we repeated the phenome-wide analyses among men (Figure S4). Additionally, our phenome-wide analyses on 43 quantitative traits, which included serum and urine biomarkers and physiological measures, identified an association of OC GRS with 13 traits. Furthermore, we found an additional 11 traits associated with at least one subtype of OC GRS (Figure S5).

We conducted two-sample MR analyses on 12 disease outcomes identified in the PheWAS. For each outcome, evidence of causality was identified for at least one type of OC (Figure 2). We found robust evidence supporting the association between a higher genetic liability of OC and an increased risk of “secondary malignant neoplasm of digestive systems” (OR 1.64, 95% CI 1.33, 2.023), “ascites” (1.48, 95% CI 1.17, 1.86), “chronic airway obstruction” (1.17, 95% CI 1.07, 1.29), and “abnormal finding on the examination of the lungs” (1.51, 95% CI 1.22, 1.87). These associations were further supported by other MR methods (Figure S6). Additionally, we observed an association between a higher genetic liability of OC and an increased risk of “ovarian cyst” (1.23, 95% CI 1.04, 1.46), as well as “lipoma including lipoma of skin and subcutaneous tissues” (1.31, 95% CI 1.06, 1.61). Notably, the weighted median and weighted mode MR methods specifically supported the association with “lipoma of the skin and subcutaneous tissues” (Figure S6). There was some evidence supporting the association between OC and higher risk of “postoperative infection” and “gram-negative septicemia”, identified in

the phenome-wide analyses on clear cell OC. However, these associations were observed only in the IVW and MR-PRESSO MR methods (Figure S6). In the MR analyses on OC subtypes (Figure 2), serous, endometrioid, and clear cell OC were all associated with “secondary malignant neoplasm of digestive systems” and “ascites”. Serous and endometrioid subtypes showed evidence of effects on the respiratory outcomes. While the case numbers were lower for mucinous OC, this histotype was the only one to show evidence of an association with “lipoma of the skin and subcutaneous tissues” as evidenced by all MR methods. However, mucinous OC was not associated with any of the other outcomes (Figure S6). The clear cell subtype was associated with “ovarian cyst”, “benign uterine neoplasms”, “uterine leiomyoma”, and “disorders of menstruation”. Both serous and clear cell subtypes increased the risk of postoperative infection traits, with clear cell also increasing risk for “gram negative septicemia” (Figure 2 and Figure S6). We conducted analyses also excluding all known OC cases, and associations with “secondary malignant neoplasm of digestive systems” and “ascites” were attenuated, while estimates for all other outcomes were generally similar to the main findings (Figure S7). The MR-Egger intercept did not suggest evidence of pleiotropy for most associations, except for the association between serous OC and “postoperative infections” ($p_{\text{intercept}} = 0.01$), which therefore should be cautiously interpreted. No evidence of pleiotropy was found using the MR-PRESSO outlier test and leave-one-out analysis (Figure S8). The estimates obtained from leave-block-out analyses were broadly consistent with the overall findings (Figure S9).

MR analyses on serum biomarkers and physiological measures supported the association of OC with lower FVC, endometrioid OC with lower oestradiol, clear cell OC with lower total cholesterol, low density lipoprotein (LDL) cholesterol, and apolipoprotein B (ApoB), and mucinous OC with higher gamma glutamyltransferase (GGT) and alanine aminotransferase (ALT) and lower creatinine levels (Figure S10). Analyses excluding all known OC cases, provided generally similar estimates to the main findings (Figure S11). For these identified associations, we did not find any evidence of pleiotropy based on the MR-Egger intercept tests ($p_{\text{intercept}} \geq 0.13$) and leave-one-out analyses. MR analyses on men provided evidence for a genetic association between mucinous OC and ALT (Figure S10).

3.2 | Shared genetic architecture

In the phenome-wide analyses assessing associations with individual OC risk-associated genetic variant, five variants were associated with at least one of 15 (11 distinct) disease outcomes (all, $p < 4.5 \times 10^{-8}$, Figure 3, Table S6), and seven

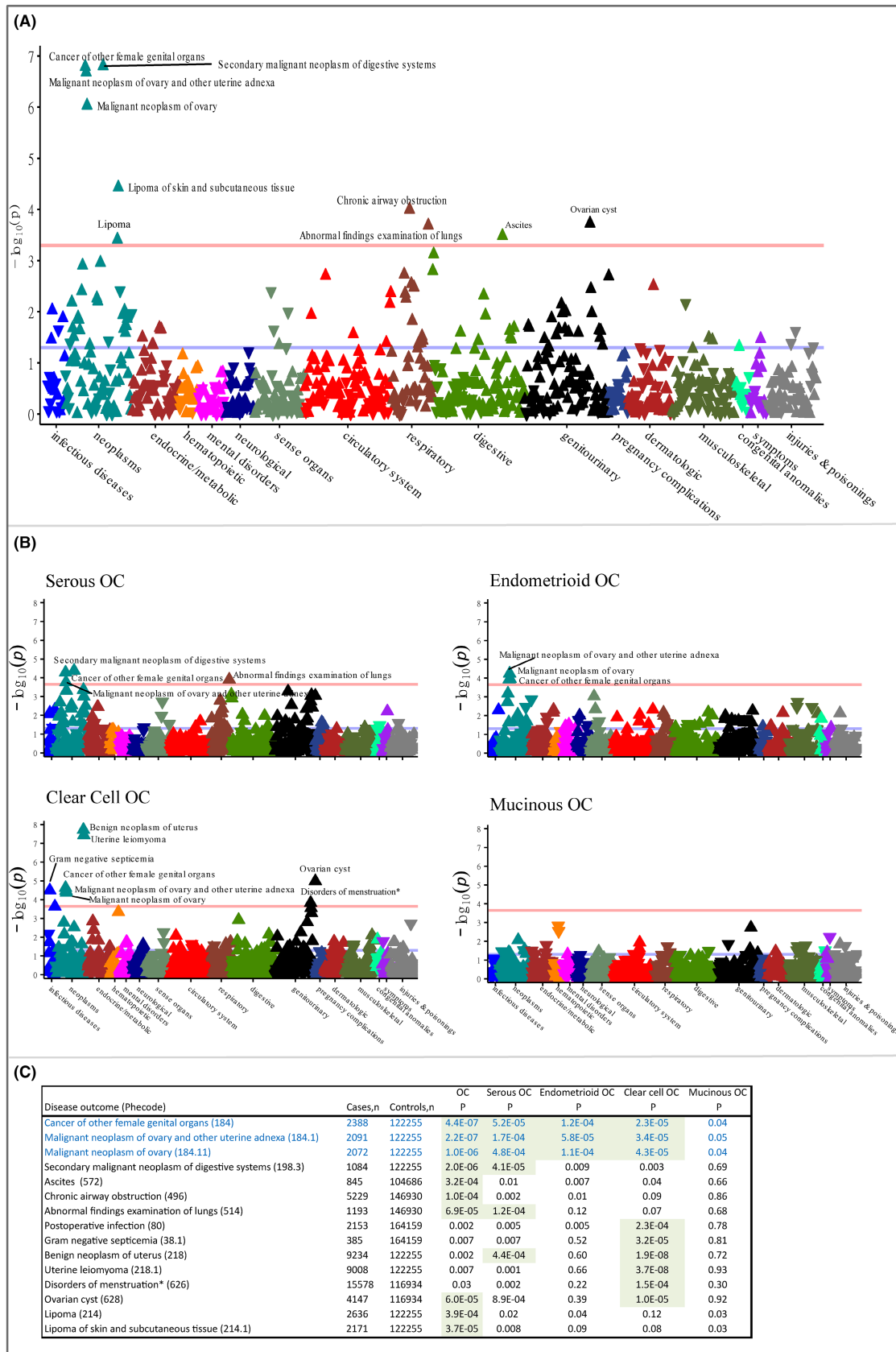


FIGURE 1 Manhattan plot showing the phenome-wide association of genetic risk scores of overall (Panel A) and subtypes (Panel B) of ovarian cancer with wide ranges of diseases. Panel C summarizes the association of the top 15 signals identified at FDR threshold ($p=0.0005$) in Panel A or Panel B across the different genetic risk scores. The association for the first three diseases (in blue) is a validation of the genetic risk scores. Green highlights indicate association that passed the FDR threshold.

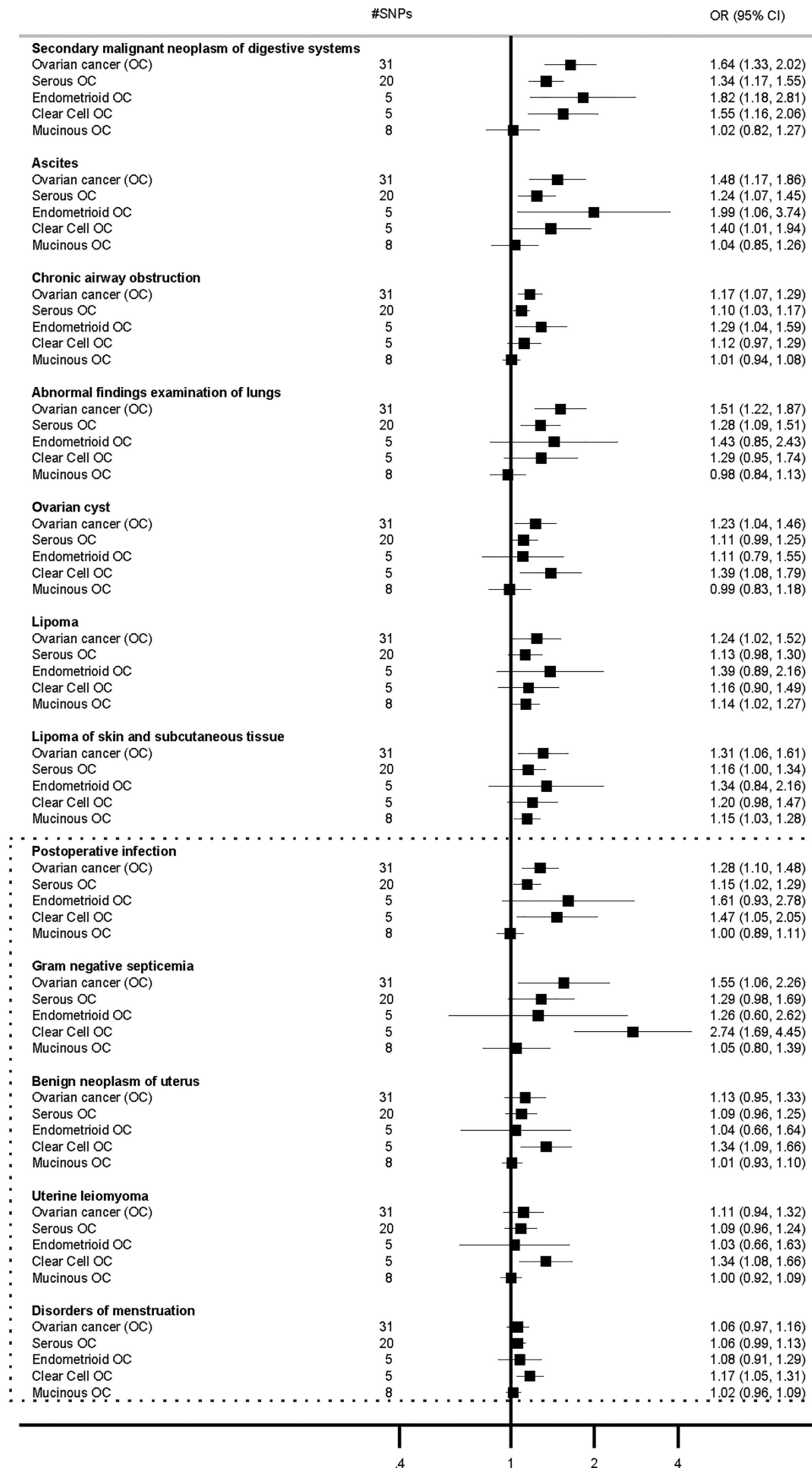


FIGURE 2 Mendelian randomization causal estimates between ovarian cancer subtypes and selected diseases outcomes. Estimates were from inverse variance weighted methods. #SNPs are number of genetic variants included in the analyses.

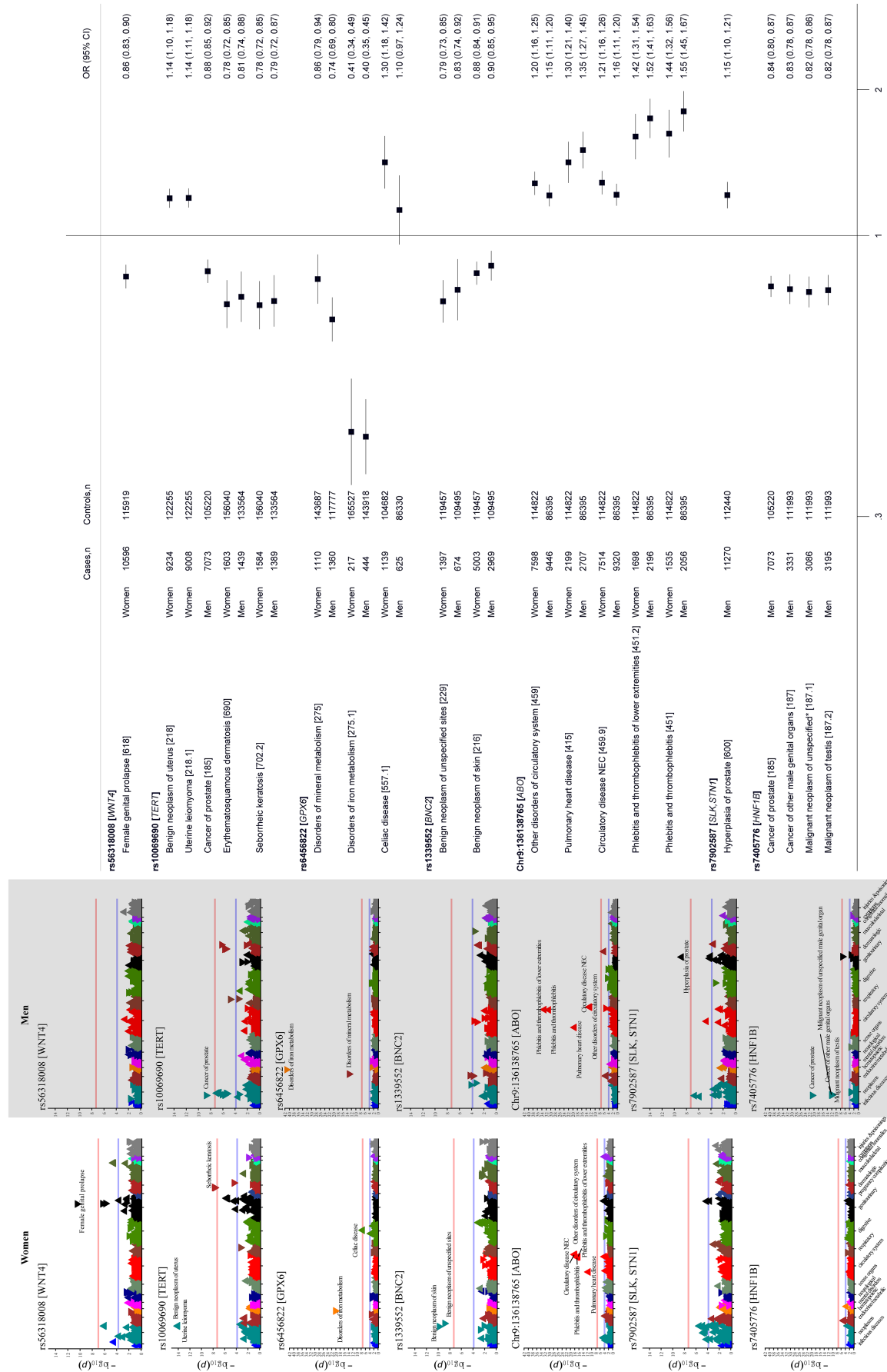


FIGURE 3 Phenome-wide association findings of ovarian cancer risk-increasing variants, with the Manhattan plot showing selected variants associations across all the diseases outcomes (with the y-axis showing the level of significance in $-\log_{10}$ scale), and the forest plot showing the selected variant—disease associations estimate among women and men.

variants were associated with at least one of 22 biomarkers and physiological measures (Figure S12). Some of these associations were also observed in the analyses among men, with selected OC variants affecting conditions of the male genital organs (Figure 3, Table S6 and Figure S12, Supplementary Results: Appendix S1). To ensure the novelty of our findings, we extensively checked the GWAS Catalog for any reports of associations with the identified (or related) diseases involving the lead or proxy variants ($r^2 \geq 0.1$). As a result, we uncovered potential novel associations between the OC risk-increasing allele from the *TERT* locus (rs10069690-T) and a lower risk of “seborrhic keratosis” (0.79, 95% CI 0.72, 0.87), in females, with similar a finding observed in males. To further validate the associations, we conducted colocalization analysis at the *TERT* locus (LD block of chr5:982252 to chr5:2132442, human genomic build of hg19) and identified rs4449583 as the candidate causal variant explaining around 70% of 0.9239 PPFC for OC-“seborrhic keratosis” (Figure 4). This variant is also strongly associated with OC in the Phelan OC GWAS¹⁵ ($p = 7.8 \times 10^{-12}$), and in LD with the index variant from this GWAS (rs7705526, $r^2 = 0.784$) and with the lead variant in the primary discovery OC GWAS²² (rs10069690, $r^2 = 0.432$, Supplementary Methods: Appendix S1). The association between the *GPX6* locus and “disorders of iron metabolism” and “celiac disease” are potentially novel, although colocalization analysis did not suggest that OC

share the same causal variants with these two diseases (Figure S13). On the contrary, we identified independent potential causal genetic variant for each “disorders of iron metabolism” (rs34409925, $r^2 = 0.207$ with rs6456822) and “celiac disease” (rs3131101, $r^2 = 0.049$ with rs6456822) in further conditional analysis (Figure S13).

4 | DISCUSSION

In this large-scale phenome-wide study, we aimed to explore comorbidity risks associated with OC, with additional analyses to identify subtype-specific associations based on genetic variants reflecting different OC histological types. We found evidence supporting a possible causal association between OC and 12 disease outcomes, including “secondary malignant neoplasm of the digestive system”, “ascites”, and respiratory disorders. We also identified some additional subtype-specific associations, including clear cell OC showing links with “gram-negative septicemia”, “uterine leiomyoma” and “disorders of menstruation”, and mucinous OC with “lipoma of the skin and subcutaneous tissue”. Analyses using 43 biomarkers also identified an association with seven biomarkers and these findings, along with those from disease analyses, may have important implications for improving the clinical management of patients with OC. Moreover,

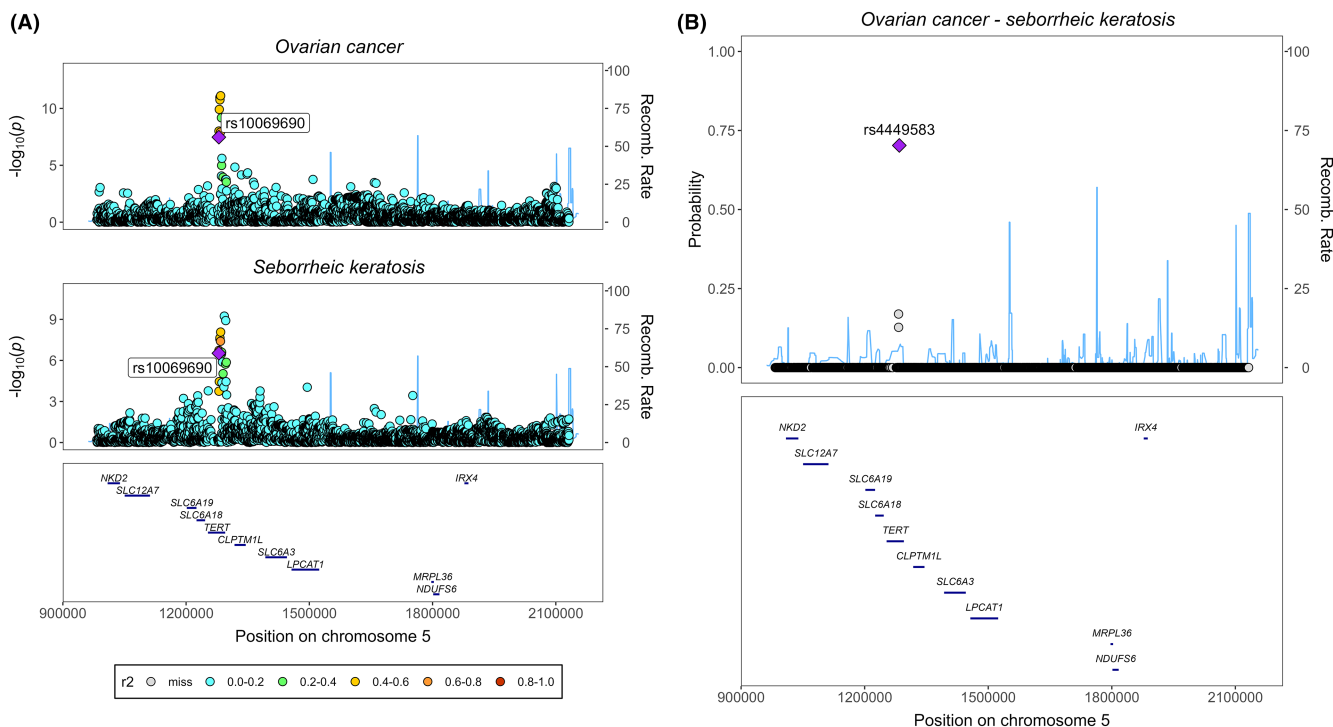


FIGURE 4 Colocalization analysis of ovarian cancer with dermatological condition at the *TERT* locus. Panel A. Stacked regional association plots of ovarian cancer and seborrhic keratosis. Panel B. Portions of the Posterior probability for full colocalization (PPFC) explained by each genetic variant at the *TERT* locus. The r^2 colour legends reflect the r^2 of each variant with respect to the lead variant.

our exploration of each OC-risk increasing variant in the PheWAS revealed a set of traits that are genetically shared with OC. This analysis not only confirmed previous GWAS findings but also shed light on novel associations deserving further investigations.

OC is the leading cause of cancer-related ascites in females.³⁴ Our MR-PheWAS study revealed genetic evidence linking ascites to OC across all subtypes except mucinous OC, which is consistent with the fact that over one-third of women with OC develop ascites during their lifetime.^{34,35} Ascites is commonly considered a sign of advanced OC,³⁶ and in line with this, the evidence for an association in our analyses was removed after excluding known OC cases. In patients with OC, the presence of ascites commonly indicates metastasis to peritoneal surfaces and reflects the extent of the disease.³⁷ Our study further established a causal association between OC and “secondary malignant neoplasms of the digestive system”, indicating intra-abdominal dissemination of the tumour. Again, this association was sensitive to the inclusion of known OC cases in our analyses, and is likely to be explained by a metastasis of the OC affecting the gastrointestinal tract or, potentially, a new synchronous gastrointestinal cancer.

OC has also been linked to respiratory complications.^{38,39} Our study provides additional supporting evidence for this association, specifically linking OC to an increased risk of “chronic airway obstruction” and “abnormal lung examinations”. The findings were further supported by spirometry measures, which showed a significant association between a higher genetic liability to OC and lower FVC. While in women with OC lowering of lung function may follow from malignant pleural effusion and tumour metastasis to the pleura³⁸ or complications related to surgical procedures,³⁹ in our analysis the association with lung disease remained when OC cases were excluded, and the spirometric association was seen both in women and in men, which may indicate a more profound biological link.

OC as reflected by our genetic study arises from the epithelium of the ovaries, endometrium, and fallopian tubes, and while most of the lung also consists of epithelium, a direct link has not been well established. Interestingly, colocalization analyses in our study also suggested that disorder of skin condition (seborrheic keratosis) shared a genetic locus (*TERT*, encoding the enzyme, telomerase) with OC, highlighting potential shared pathomechanisms involving the epithelium, and implicating telomerase function in the opposing risks of OC and the skin disorder (the allele that increase OC linked with reduced risk of seborrheic keratosis).^{40,41}

Debulking cytoreductive surgery is the standard procedure for advanced OC. It involves the surgical removal of female reproductive organs such as the ovaries, fallopian

tubes, uterus, as well as omentum and organs of the digestive system, such as the bowel, depending on the stage of OC.⁴² The invasive nature of this intervention,⁴² combined with compromised immune system function resulting from the disease,⁴³ may explain the observed association between OC and “postoperative infection” and “gram-negative septicemia”. It is important to note that these associations were predominantly observed in clear cell OC, which has also been associated with “benign neoplasms of the uterus”, “uterine leiomyoma”, and “disorders of menstruation”. The presence of these conditions, which primarily involve the uterine tissue, could reflect the well-established association between endometriosis and clear cell OC.^{44,45} Our study also identified OC subtype-specific associations between mucinous OC and “lipoma of the skin and subcutaneous tissue”. Regarding the biomarkers, endometrioid OC was associated with lower serum oestradiol level, clear cell OC was associated with lower serum total cholesterol, LDL cholesterol, and ApoB, while mucinous OC was associated with higher serum GGT and ALT as well as lower creatinine levels. These findings suggest that comorbidities and selected biomarkers may play a role in identifying and characterizing different subtypes of OC.

Beyond the causal link between OC and disease outcomes and biomarkers, genetic variants contributing to a higher risk of OC may also be associated with other diseases and biomarkers. The *ABO* locus is a well-known pleiotropic locus with established associations with cardiovascular and cerebrovascular traits.^{46–51} Our study confirmed these existing findings, as we also observed associations between the *ABO* variant and conditions such as “pulmonary heart disease”, “thrombophlebitis”, and “other disorders of circulatory system”. Additionally, we confirmed the association between *ABO* with many cardiometabolic biomarkers, including associations with markers of glucose and liver metabolism, confirming the highly pleiotropic nature of this variant.⁵²

Our study identified a significant association between the *GPX6* variant (rs6456822) and a lower risk of disorders related to iron metabolism, which is a critical component of overall health.⁵³ However, while the lead variant rs6456822 showed this association, our colocalization analysis identified an independent potential causal variant, rs34409925, for “disorders of iron metabolism”. We also found another potential causal variant at the *GPX6* locus, rs3131101, for “celiac disease”. The OC lead variant rs6456822 was also associated with total protein levels in women, as well as with SHBG levels in men. Our study also identified a potential novel association between the *TERT* locus (rs10069690) and “seborrheic keratosis”. The follow up colocalization analysis strengthened the shared genetic structures between this skin condition and OC, further suggesting that rs444583

may be a causal variant. Importantly, we confirmed previous associations of the *TERT* variant with uterine leiomyoma^{45,54} and total protein levels in women,^{50,52} as well as with prostate cancer in men.⁵⁵ Our findings highlight the value of using PheWAS for validating previous GWAS findings and identifying potential novel variants for multiple traits including diseases and biomarkers that never been tested before.⁵⁶

Our study has a number of strengths. Firstly, it is the largest ($n=181,203$) and most comprehensive study to date that examines the health risks associated with a higher genetic susceptibility to OC. Secondly, by including OC subtypes in our analysis, our study can identify subtype-specific comorbidities, supporting more effective treatment and management. Lastly, our study extended the phenome-wide analyses to each OC-risk-increasing variant. This approach enables us to confirm previously identified variant-disease associations and also identify potential novel associations, thus providing a more comprehensive understanding of the shared genetic links between OC and other diseases. Further incorporation of colocalization analysis is another strength of this study. It provided evidence of a shared genetic locus (*TERT*) between OC and “seborrhic keratosis”, and identified a potential novel causal variant, rs4449583, for this skin condition.

Our study also has some limitations. To ensure the validity of MR, the genetic instrument used must be associated with the risk factor (in our case, OC), not associated with the confounders of the risk factor-outcome association, and not influence the outcome except through its association with the risk factor. While we used a genetic instrument that associates with OC and conducted our analysis on a white-European population while adjusting for confounding factors, we cannot rule out the absence of horizontal pleiotropic effects, which will affect interpretation and can bias the MR estimates.⁵⁷ To address this concern, we repeated our analysis in men to further verify whether the association we observed is genuinely related to OC or a reflection of genetic variants indicating shared mechanistic pathways between both men and women. Another potential limitation of our study is that the UK Biobank participants are relatively healthy compared to the general population,⁵⁸ introducing to collider bias due to the “healthy volunteer” selection bias.⁵⁹ However, the modest collider bias in the context of MR is less likely to affect MR results significantly when the selection into the study was not strongly influenced by the risk factor.⁵⁹ Furthermore, the presence of biologically plausible causal associations in our study provides support for the validity of our results despite this limitation. It could be noted as a limitation of our study that our investigation into comorbidities associated with OC was primarily conducted in a population predominantly not diagnosed with OC.

However, sensitivity analyses excluding all known OC cases confirmed that the associations with OC complications relating to ascites and gastrointestinal metastases are related to the disease per se, while the other observed disease and trait associations may reflect earlier stages of OC development or other underlying biological differences. Since our studies are limited to the white European population, the generalizability of our findings to other populations is uncertain. Future studies involving diverse populations are warranted.

In conclusion, our study identified robust associations between OC and comorbidities related to the digestive and respiratory systems. Furthermore, we observed specific robust associations for different subtypes of OC, such as “gram-negative septicemia”, “uterine leiomyoma”, and “disorders of menstruation” for clear cell OC, and lipoma for mucinous OC. Interestingly, our study points to a shared genetic architecture between OC and a skin condition, leading to a discovery of a novel variant which merits further investigation.

AUTHOR CONTRIBUTIONS

Anwar Mulugeta: Conceptualization (equal); data curation (equal); formal analysis (lead); investigation (lead); methodology (lead); project administration (lead); validation (equal); visualization (lead); writing – original draft (lead); writing – review and editing (lead). **Amanda Lumsden L:** Project administration (supporting); validation (supporting); writing – review and editing (supporting). **Iqbal Madakkattel:** Validation (supporting); writing – review and editing (supporting). **David Stacey:** Formal analysis (supporting); methodology (supporting); validation (supporting); writing – review and editing (supporting). **S. Hong Lee:** Validation (supporting); writing – review and editing (supporting). **Johanna Mäenpää:** Validation (supporting); writing – review and editing (supporting). **Martin Oehler:** Validation (supporting); writing – review and editing (supporting). **Elina Hyppönen:** Conceptualization (equal); data curation (equal); funding acquisition (lead); methodology (equal); project administration (equal); resources (equal); supervision (equal); validation (equal); writing – review and editing (equal).

ACKNOWLEDGMENTS

The authors are very grateful to the UK Biobank and UK Biobank Participants. Open access publishing facilitated by University of South Australia, as part of the Wiley - University of South Australia agreement via the Council of Australian University Librarians.

FUNDING INFORMATION

This research was funded by Medical Research Future Fund of Australia (MRF2007431).

CONFLICT OF INTEREST STATEMENT

The authors declare that they have no competing interests.

DATA AVAILABILITY STATEMENT

This research has been conducted using the UK Biobank Resource under Application Number 89630. The data and analytical code that support the findings of this study are available on application to the UK Biobank (<https://www.ukbiobank.ac.uk/enable-your-research/apply-for-access>).

ETHICS STATEMENT

The study was conducted under UK Biobank application number 89630. UK Biobank study has ethical approval from the National Information Governance Board for Health and Social Care and Northwest Multicenter Research Ethics Committee (11/NW/0382).

CONSENT TO PARTICIPATE

UK Biobank obtained informed consent from all participants to collect and use linked data for future research use.

ORCID

Anwar Mulugeta  <https://orcid.org/0000-0002-8018-3454>

REFERENCES

- DeSantis CE, Lin CC, Mariotto AB, et al. Cancer treatment and survivorship statistics, 2014. *CA Cancer J Clin*. 2014;64:252-271.
- Kurman RJ, Shih IM. The origin and pathogenesis of epithelial ovarian cancer: a proposed unifying theory. *Am J Surg Pathol*. 2010;34:433-443.
- Sankaranarayanan R, Ferlay J. Worldwide burden of gynaecological cancer: the size of the problem. *Best Pract Res Clin Obstet Gynaecol*. 2006;20:207-225.
- Perets R, Wyant GA, Muto KW, et al. Transformation of the fallopian tube secretory epithelium leads to high-grade serous ovarian cancer in Brca;Tp53;Pten models. *Cancer Cell*. 2013;24:751-765.
- Lheureux S, Gourley C, Vergote I, Oza AM. Epithelial ovarian cancer. *Lancet*. 2019;393:1240-1253.
- Zhang S, Cheng C, Lin Z, et al. The global burden and associated factors of ovarian cancer in 1990-2019: findings from the global burden of disease study 2019. *BMC Public Health*. 2022;22:1455.
- Jiao YS, Gong TT, Wang YL, Wu QJ. Comorbidity and survival among women with ovarian cancer: evidence from prospective studies. *Sci Rep*. 2015;5:11720.
- Lee L, Cheung WY, Atkinson E, Krzyzanowska MK. Impact of comorbidity on chemotherapy use and outcomes in solid tumors: a systematic review. *J Clin Oncol*. 2011;29:106-117.
- Bush WS, Oetjens MT, Crawford DC. Unravelling the human genome-phenome relationship using phenome-wide association studies. *Nat Rev Genet*. 2016;17:129-145.
- Davies NM, Holmes MV, Davey SG. Reading mendelian randomisation studies: a guide, glossary, and checklist for clinicians. *BMJ*. 2018;362:k601.
- Sudlow C, Gallacher J, Allen N, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med*. 2015;12:e1001779.
- Wei WQ, Bastarache LA, Carroll RJ, et al. Evaluating phenocodes, clinical classification software, and ICD-9-CM codes for phenome-wide association studies in the electronic health record. *PLoS One*. 2017;12:e0175508.
- King C, Mulugeta A, Nabi F, Walton R, Zhou A, Hyppönen E. Mendelian randomization case-control PheWAS in UK biobank shows evidence of causality for smoking intensity in 28 distinct clinical conditions. *EClinicalMedicine*. 2020;26:100488.
- Verma A, Bradford Y, Dudek S, et al. A simulation study investigating power estimates in phenome-wide association studies. *BMC Bioinformatics*. 2018;19:120.
- Phelan CM, Kuchenbaecker KB, Tyrer JP, et al. Identification of 12 new susceptibility loci for different histotypes of epithelial ovarian cancer. *Nat Genet*. 2017;49:680-691.
- Song H, Ramus SJ, Tyrer J, et al. A genome-wide association study identifies a new ovarian cancer susceptibility locus on 9p22.2. *Nat Genet*. 2009;41:996-1000.
- Goode EL, Chenevix-Trench G, Song H, et al. A genome-wide association study identifies susceptibility loci for ovarian cancer at 2q31 and 8q24. *Nat Genet*. 2010;42:874-879.
- Bolton KL, Tyrer J, Song H, et al. Common variants at 19p13 are associated with susceptibility to ovarian cancer. *Nat Genet*. 2010;42:880-884.
- Rafnar T, Gudbjartsson DF, Sulem P, et al. Mutations in BRIP1 confer high risk of ovarian cancer. *Nat Genet*. 2011;43:1104-1107.
- Couch FJ, Wang X, McGuffog L, et al. Genome-wide association study in BRCA1 mutation carriers identifies novel loci associated with breast and ovarian cancer risk. *PLoS Genet*. 2013;9:e1003212.
- Pharoah PD, Tsai YY, Ramus SJ, et al. GWAS meta-analysis and replication identifies three new susceptibility loci for ovarian cancer. *Nat Genet*. 2013;45:362-370.
- Kuchenbaecker KB, Ramus SJ, Tyrer J, et al. Identification of six new susceptibility loci for invasive epithelial ovarian cancer. *Nat Genet*. 2015;47:164-171.
- Kelemen LE, Lawrenson K, Tyrer J, et al. Genome-wide significant risk associations for mucinous ovarian carcinoma. *Nat Genet*. 2015;47:888-897.
- Bycroft C, Freeman C, Petkova D, et al. The UK biobank resource with deep phenotyping and genomic data. *Nature*. 2018;562:203-209.
- Carroll RJ, Bastarache L, Denny JC. R PheWAS: data analysis and plotting tools for phenome-wide association studies in the R environment. *Bioinformatics*. 2014;30:2375-2376.
- Bowden J, Del Greco MF, Minelli C, Davey Smith G, Sheehan N, Thompson J. A framework for the investigation of pleiotropy in two-sample summary data mendelian randomization. *Stat Med*. 2017;36:1783-1802.
- Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent estimation in mendelian randomization with some invalid instruments using a weighted median estimator. *Genet Epidemiol*. 2016;40:304-314.
- Hartwig FP, Davey Smith G, Bowden J. Robust inference in summary data mendelian randomization via the zero modal pleiotropy assumption. *Int J Epidemiol*. 2017;46:1985-1998.

29. Verbanck M, Chen CY, Neale B, Do R. Detection of widespread horizontal pleiotropy in causal relationships inferred from mendelian randomization between complex traits and diseases. *Nat Genet.* 2018;50:693-698.
30. Bowden J, Del Greco MF, Minelli C, Davey Smith G, Sheehan NA, Thompson JR. Assessing the suitability of summary data for two-sample mendelian randomization analyses using MR-egger regression: the role of the I^2 statistic. *Int J Epidemiol.* 2016;45:1961-1974.
31. Skrivankova VW, Richmond RC, Woolf BAR, et al. Strengthening the reporting of observational studies in epidemiology using mendelian randomisation (STROBE-MR): explanation and elaboration. *BMJ.* 2021;375:n2233.
32. Foley CN, Staley JR, Breen PG, et al. A fast and efficient colocalization algorithm for identifying shared genetic risk factors across multiple traits. *Nat Commun.* 2021;12:764.
33. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience.* 2015;4:7.
34. Parsons SL, Lang MW, Steele RJC. Malignant ascites: a 2-year review from a teaching hospital. *European Journal of Surgical Oncology (EJSO).* 1996;22:237-239.
35. Ayantunde AA, Parsons SL. Pattern and prognostic factors in patients with malignant ascites: a retrospective study. *Ann Oncol.* 2007;18:945-949.
36. Kipps E, Tan DS, Kaye SB. Meeting the challenge of ascites in ovarian cancer: new avenues for therapy and research. *Nat Rev Cancer.* 2013;13:273-282.
37. Ahmed N, Stenvers KL. Getting to know ovarian cancer ascites: opportunities for targeted therapy-based translational research. *Front Oncol.* 2013;3:256.
38. Porcel JM, Diaz JP, Chi DS. Clinical implications of pleural effusions in ovarian cancer. *Respirology.* 2012;17:1060-1067.
39. Xu M, Zhang W, Gao C, Zhou Y, Xie Y. Postoperative pulmonary complications and outcomes in cytoreductive surgery for ovarian cancer: a propensity-matched analysis. *BMC Anesthesiol.* 2022;22:120.
40. Buckingham EM, Klingelhut AJ. The role of telomeres in the ageing of human skin. *Exp Dermatol.* 2011;20:297-302.
41. Chen B, Yan Y, Wang H, Xu J. Association between genetically determined telomere length and health-related outcomes: a systematic review and meta-analysis of mendelian randomization studies. *Aging Cell.* 2023;22:e13874.
42. Pomel C, Jeyarajah A, Oram D, et al. Cytoreductive surgery in ovarian cancer. *Cancer Imaging.* 2007;7:210-215.
43. Cubillos-Ruiz Juan R, Silberman Pedro C, Rutkowski Melanie R, et al. ER stress sensor XBP1 controls anti-tumor immunity by disrupting dendritic cell homeostasis. *Cell.* 2015;161:1527-1538.
44. Mortlock S, Corona RI, Kho PF, et al. A multi-level investigation of the genetic relationship between endometriosis and ovarian cancer histotypes. *Cell Rep Med.* 2022;3:100542.
45. Gallagher CS, Mäkinen N, Harris HR, et al. Genome-wide association and epidemiological analyses reveal common genetic origins between uterine leiomyomata and endometriosis. *Nat Commun.* 2019;10:4857.
46. Schunkert H, König IR, Kathiresan S, et al. Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nat Genet.* 2011;43:333-338.
47. Lindström S, Brody JA, Turman C, et al. A large-scale exome array analysis of venous thromboembolism. *Genet Epidemiol.* 2019;43:449-457.
48. Klarin D, Busenkell E, Judy R, et al. Genome-wide association analysis of venous thromboembolism identifies new risk loci and genetic overlap with arterial vascular disease. *Nat Genet.* 2019;51:1574-1579.
49. Shah S, Henry A, Roselli C, et al. Genome-wide association and mendelian randomisation analysis provide insights into the pathogenesis of heart failure. *Nat Commun.* 2020;11:163.
50. Sakaue S, Kanai M, Tanigawa Y, et al. A cross-population atlas of genetic associations for 220 human phenotypes. *Nat Genet.* 2021;53:1415-1424.
51. Klarin D, Damrauer SM, Cho K, et al. Genetics of blood lipids among ~300,000 multi-ethnic participants of the million veteran program. *Nat Genet.* 2018;50:1514-1523.
52. Sinnott-Armstrong N, Tanigawa Y, Amar D, et al. Genetics of 35 blood and urine biomarkers in the UK biobank. *Nat Genet.* 2021;53:185-194.
53. Dev S, Babbitt JL. Overview of iron metabolism in health and disease. *Hemodialysis International Symposium on Home Hemodialysis.* 2017;21(Suppl 1):S6-S20.
54. Rafnar T, Gunnarsson B, Stefansson OA, et al. Variants associating with uterine leiomyoma highlight genetic background shared by various cancers and hormone-related traits. *Nat Commun.* 2018;9:3636.
55. Schumacher FR, Al Olama AA, Berndt SI, et al. Association analyses of more than 140,000 men identify 63 new prostate cancer susceptibility loci. *Nat Genet.* 2018;50:928-936.
56. Pendergrass SA, Brown-Gentry K, Dudek SM, et al. The use of phenome-wide association studies (PheWAS) for exploration of novel genotype-phenotype relationships and pleiotropy discovery. *Genet Epidemiol.* 2011;35:410-422.
57. Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through egger regression. *Int J Epidemiol.* 2015;44:512-525.
58. Fry A, Littlejohns TJ, Sudlow C, et al. Comparison of socio-demographic and health-related characteristics of UK biobank participants with those of the general population. *Am J Epidemiol.* 2017;186:1026-1034.
59. Gkatzionis A, Burgess S. Contextualizing selection bias in mendelian randomization: how bad is it likely to be? *Int J Epidemiol.* 2019;48:691-701.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Mulugeta A, Lumsden AL, Madakkat I, et al. Phenome-wide association study of ovarian cancer identifies common comorbidities and reveals shared genetics with complex diseases and biomarkers. *Cancer Med.* 2024;13:e7051. doi:[10.1002/cam4.7051](https://doi.org/10.1002/cam4.7051)