International Journal of Women's Health

Dovepress
Taylor & Francis Group

Open Access Full Text Article

ORIGINAL RESEARCH

Mendelian Randomization to Examine the Causal Effects of Cystatin on Ovarian Lesions

Tianqing Yan[1], Lin Guo[1,2], Renquan Lu[1,2]

[1]Department of Clinical Laboratory, Fudan University Shanghai Cancer Center, Shanghai, 200032, People's Republic of China; [2]Department of Oncology, Shanghai Medical College, Fudan University, Shanghai, 200032, People's Republic of China

Correspondence: Renquan Lu, Email renquanlu@fudan.edu.cn

Background: Cystatin, a superfamily of cysteine protease inhibitors, are implicated in extracellular matrix remodeling, immune modulation, and tumor progression. Observational studies have reported associations between specific cystatin and gynecological pathologies, including ovarian cancer. However, whether these associations reflect causality remains uncertain due to potential confounding and reverse causation.

Methods: We performed a two-sample Mendelian randomization (MR) study to investigate the causal relationships between genetically predicted levels of seven cystatin subtypes (Cystatin B, C, D, F, M, S, and Cystatin 8) and the risk of various ovarian lesions. Genetic instruments (single nucleotide polymorphisms, SNPs) for cystatin were selected from genome-wide association studies (GWAS) at a significance threshold of $P < 5\times10^{-6}$, with clumping for linkage disequilibrium (LD $R^2 < 0.001$, window = 10,000 kb). Summary-level data for outcomes including ovarian cysts, primary ovarian failure, ovarian torsion, and major histologic subtypes of ovarian cancer (high/low-grade serous, mucinous, clear cell, endometrioid) were obtained from the FinnGen, UK Biobank and Ovarian Cancer Association Consortium (OCAC). The primary analysis used the inverse-variance weighted (IVW) method, complemented by Weighted Median, MR-Egger, Mode-based methods, and sensitivity analyses for pleiotropy and heterogeneity.

Results: Genetically predicted Cystatin-8 was significantly associated with a reduced risk of endometrioid ovarian cancer (OR = 0.898, 95% CI: 0.836–0.965, P=0.003). Cystatin-C showed a suggestive protective effect against high-grade serous ovarian cancer (OR = 0.908, 95% CI: 0.828–0.994, P=0.038). Cystatin-F was associated with a reduced risk of polycystic ovarian syndrome (OR = 0.852, 95% CI: 0.725–0.999, P = 0.049). No significant causal relationships were observed for the other cystatin subtypes with the studied ovarian lesions. Sensitivity analyses were conducted to validate the MR assumptions. Cochran's Q test was used to assess heterogeneity across the genetic instruments. The MR-Egger intercept test was applied to detect and adjust for directional horizontal pleiotropy. Additionally, a leave-one-out analysis was performed to ensure that no single SNP disproportionately drove the causal estimates.

Conclusion: This MR study provides genetic evidence supporting a potential causal, protective role for Cystatin-8 against endometrioid ovarian cancer, and suggestive protective roles for Cystatin-C in high-grade serous ovarian cancer and Cystatin-F in polycystic ovarian syndrome. These findings highlight specific cystatin as potential biomarkers or etiological factors warranting further mechanistic and clinical investigation in ovarian pathophysiology.

Keywords: cysteine protease inhibitors, ovarian cancer, ovarian lesions, Mendelian randomization, causal inference

Introduction

Ovarian lesions encompass a spectrum of conditions from benign disorders, such as ovarian cysts and polycystic ovarian syndrome (PCOS), to malignant neoplasms, which collectively represent a significant cause of morbidity and mortality in women globally.[1] Epithelial ovarian cancer (EOC) is the most lethal gynecologic malignancy, with distinct histologic subtypes including high-grade serous (HGSOC), low-grade serous (LGSOC), clear cell (OCCC), endometrioid (ENOC), and mucinous (MOC), exhibiting diverse molecular profiles, clinical behaviors, and prognoses.[2] Despite advances in treatment, the 5-year survival rate for advanced-stage disease remains below 50%, underscoring the urgent need for a deeper understanding of etiological factors and early biomarkers.[3]

https://doi.org/10.2147/IJWH.S605528
Received: 26 February 2026
Accepted: 11 May 2026
Published: 15 May 2026

The cystatin superfamily comprises endogenous inhibitors of cysteine cathepsins, a class of lysosomal proteases involved in protein turnover, extracellular matrix (ECM) degradation, antigen presentation, and apoptosis.[4] By regulating cathepsin activity, cystatin play crucial roles in maintaining tissue homeostasis, modulating inflammatory responses, and influencing tumor microenvironments. Dysregulation of the cystatin-cathepsin axis has been implicated in various cancers, including breast, colorectal, and lung malignancies, where it can affect invasion, metastasis, and immune evasion.[5,6] In ovarian pathology, alterations in cystatin expression have been identified, with certain cystatins showing potential links to disease progression.[7] However, findings from conventional observational studies in this field are often constrained by methodological limitations, including confounding from environmental and lifestyle factors, as well as the possibility of reverse causality where disease status may itself influence biomarker concentrations. Consequently, establishing a clear causal relationship between circulating cystatin levels and ovarian disease risk remains challenging within the framework of traditional epidemiological approaches.

Mendelian randomization (MR) has emerged as a powerful method for causal inference in observational data. By using genetic variants as instrumental variables (IVs) for modifiable exposures, MR mimics the random allocation of a randomized controlled trial, largely avoiding confounding and reverse causation.[8] The approach relies on three core assumptions:[9] 1) the genetic instruments are robustly associated with the exposure; 2) they are independent of confounders; and 3) they affect the outcome only through the exposure (no horizontal pleiotropy). Two-sample MR, which utilizes summary-level data from independent genome-wide association studies (GWAS) for exposure and outcome, has been successfully applied to elucidate causal relationships in oncology,[10,11] as demonstrated in recent studies exploring galectins in gynecologic cancers[12] and cathepsins in ovarian tumors.[13]

Inspired by the analytical frameworks of these studies and given the biological plausibility of cystatin involvement in ovarian pathophysiology, we conducted a comprehensive two-sample MR analysis. This study aims to investigate the potential causal effects of seven genetically predicted cystatin subtypes on a broad range of ovarian lesions, from benign conditions to malignant histological subtypes. Our findings aim to clarify the etiological roles of cystatin and identify potential targets for risk stratification and therapeutic intervention.

Methods

Study Design and MR Assumptions

The study examined the causal effects of seven cystatin subtypes (exposures) on various ovarian lesions (outcomes), including benign conditions and malignant neoplasms, while accounting for potential confounders and strictly conformed to the proposed STROBE-MR Statement in the analysis[14] (Figure 1). The MR analysis was predicated on three fundamental assumptions:[9] 1) Relevance: The selected single nucleotide polymorphisms (SNPs) are strongly associated with circulating levels of the respective cystatin. 2) Independence: These SNPs are not associated with any known or unknown confounders of the cystatin-ovarian lesion relationship. 3) Exclusivity: The SNPs influence the risk of ovarian lesions only via their effect on cystatin levels, not through other biological pathways (no horizontal pleiotropy).

Data Sources for Exposures (Cystatin) and Outcomes (Ovarian Lesions)

Summary statistics data for SNPs related to plasma Cystatins were extracted from the most comprehensive publicly available genome-wide association studys (GWAS) summary statistics to date (MRC-Integrative Epidemiology Unit (IEU) OpenGWAS database, https://opengwas.io/, accessed on 2025.12.15). The Cystatin-B and Cystatin-C datasets includes 3394 individuals and 437,846 individuals of European ancestry respectively.[15,16] Genetic data for the Cystatin-S include a European-ancestry population, covering 501,428 SNPs.[17] The datasets for Cystatin-8, Cystatin-D, Cystatin-F, Cystatin-M comprise 3301 individuals of European ancestry.[18] Data on ovarian cancer (25,509 cases among 66,450 individuals), HGSOC (13,037 cases among 53,978 individuals), LGSOC (1012 cases among 41,953 individuals), MOC (1417 cases among 42,358 individuals), OCCC (1366 cases among 42,307 individuals), and ENOC (2810 cases among 43,751 individuals) were sourced from the Ovarian Cancer Association Consortium (OCAC).[19] The data on ovarian cysts comprise a European-ancestry sample of 462,933 individuals (3877 cases and 459,056 controls) sourced from the UK Biobank via the MRC-IEU consortium. Data on primary ovarian failure (254 cases, 118,228 controls), torsion of ovary/ovarian pedicle/

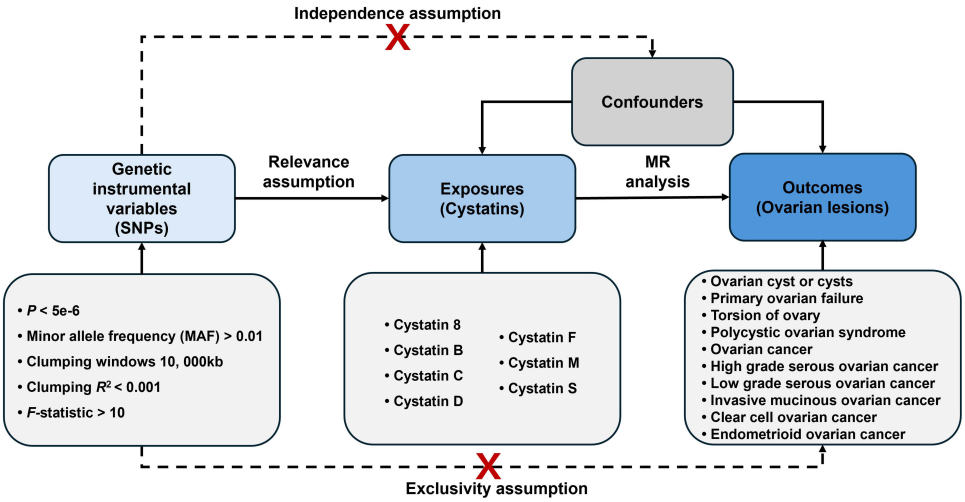

Figure 1 Study design of the Mendelian randomization analysis. Genetic instrumental variables (IVs) for cystatin subtypes were selected as SNPs associated with each exposure at a genome-wide significance threshold ($P < 5\times10^{-6}$). Clumping was performed with a window of 10,000 kb and a linkage disequilibrium (LD) threshold of $R^2 < 0.001$. Only instruments with an *F*-statistic > 10 were retained to ensure strong instrumental strength. The red crosses over the dotted lines indicate that potential confounders are blocked or adjusted for in the MR design.

fallopian tube (318 cases, 68,969 controls), and polycystic ovarian syndrome (642 cases, 118,228 controls) were obtained from the FinnGen consortium (www.finnbb.fi). The characteristics of each cystatin dataset, and outcome datasets are detailed in Table 1. Ethical approval and participant informed consent were obtained in all original GWAS. The current analysis utilized fully anonymized summary-level data; therefore, no additional ethical review was required for this study.

Table 1 Description of Exposures Traits and Outcomes (European)

Phenotype	Sample Size (Case)	SNPs	GWAS ID	Consortium
Exposures				
Cystatin-8	3301	10,534,735	prot-a-706	NA
Cystatin-B	3394	5,270,646	prot-b-3	NA
Cystatin-C	437,846	4,232,088	ebi-a-GCST90025945	NA
Cystatin-D	3301	10,534,735	prot-a-702	NA
Cystatin-F	3301	10,534,735	prot-a-705	NA
Cystatin-M	3301	10,534,735	prot-a-704	NA
Cystatin-S	NA	501,428	prot-c-3802_50_1	NA
Outcomes				
Ovarian cyst or cysts	462,933	9,851,867	ukb-b-15025	MRC-IEU
Primary ovarian failure	NA	16,379,677	finn-b-E4_OVARFAIL	NA
Torsion of ovary	NA	16,376,184	finn-b-N14_OVARTORS	NA
PCOS	NA	16,379,676	finn-b-E4_POCS	NA
OC	66,450 (25,509)	NA	ieu-a-1120	OCAC
HGSC	53,978 (13,037)	NA	ieu-a-1121	OCAC
LGSC	41,953 (1012)	NA	ieu-a-1122	OCAC
MOC	42,358 (1417)	NA	ieu-a-1123	OCAC
OCCC	42,307 (1366)	NA	ieu-a-1124	OCAC
ENOC	43,751 (2810)	NA	ieu-a-1125	OCAC

Abbreviations: GWAS, genome-wide association studies; PCOS, polycystic ovarian syndrome; OC, ovarian cancer; HGSOC, high-grade serous ovarian cancer; LGSOC, low-grade serous ovarian cancer; MOC, mucinous ovarian cancer; OCCC, ovarian clear cell cancer; ENOC, endometrioid ovarian cancer; OCAC, ovarian cancer association consortium; NA, not available.

https://doi.org/10.2147/IJWH.S605528

Selection of Genetic Instrumental Variables (IVs)

The SNPs independently associated with its circulating levels at a genome-wide significance threshold of $P < 5\times10^{-6}$ were selected as candidate IVs. A less stringent threshold than the conventional 5×10^{-8} was used to ensure sufficient numbers of instruments for MR analysis, a common practice in protein quantitative trait locus (pQTL) studies.[14] SNPs with a minor allele frequency (MAF) > 0.01 were selected from the analysis. We then performed LD clumping (R^2<0.001, window size = 10,000 kb) to ensure the independence of selected SNPs.[14] Palindromic SNPs with ambiguous strand orientation were removed. For each SNP, we calculated the proportion of variance explained (R^2) and the *F*-statistic to assess instrument strength, retaining only instruments with an *F*-statistic > 10 to mitigate weak instrument bias.[20] The *F*-statistic was calculated using the following formula: $F = [R^2 \times (N\text{-}1\text{-}K)] / [K \times (1\text{-}R^2)]$. The variance explained (R^2) was derived from the formula: $R^2 = 2 \times MAF \times (1\text{-}MAF) \times \beta^2$, where β represents the estimated effect size of the exposure.

Statistical Analysis of Mendelian Randomization

The primary causal estimates were calculated using the inverse-variance weighted (IVW) method with multiplicative random effects, which provides a meta-analysis of Wald ratios for each SNP.[21,22] To ensure robustness and account for potential violations of MR assumptions, four supplementary methods were employed. The Weighted Median remains consistent despite up to 50% invalid instruments,[23] whereas MR-Egger regression tests and adjusts for pleiotropy via an intercept term.[20] The Weighted Mode clusters SNPs by estimate similarity, in comparison to the more basic Simple Mode estimator. Results for associations are presented as odds ratios (ORs) with 95% confidence intervals (CIs) per unit increase in genetically predicted cystatin level.

Sensitivity Analyses

We conducted extensive sensitivity analyses to validate the primary results. Cochran's Q statistic was used to evaluate heterogeneity among the Wald ratios from individual SNPs. A significant Q statistic ($P < 0.05$) indicates potential heterogeneity, possibly due to pleiotropy. A random-effects IVW model was applied in the presence of significant heterogeneity; conversely, a fixed-effects IVW model was used when no such heterogeneity was detected. The MR-Egger intercept test was performed to assess the horizontal pleiotropy. A non-zero intercept ($P < 0.05$) suggests the presence of overall directional pleiotropy. Leave-One-Out Analysis was applied to determine if the causal estimate was driven by a single influential variant. All statistical analyses were performed using R software (version 4.5.2) with the TwoSampleMR (version 0.6.29) packages. Visual representations, including scatter plots, funnel plots, and forest plots, were employed to illustrate the analytical results. The scatter plot depicts the effect relationship of IVs on both the exposure and outcome, whereas the funnel plot assesses potential bias, and the forest plot displays the effect estimates of individual SNPs along with their consistency.

Results

Characteristics of Genetic Instruments and Causal Effects of Cystatin on Ovarian Lesions

The selection process yielded strong instrumental variables for each cystatin subtype, with *F*-statistics all exceeding the threshold of 10, indicating a low risk of weak instrument bias. The number of SNPs used for each cystatin-outcome pair analysis varied and is reported in the forest plots. The IVW MR estimates for all cystatin-ovarian lesion pairs are summarized in the comprehensive forest plot (Figure 2), where statistically significant protective effects of genetically predicted higher Cystatin-8 levels on the risk of endometrioid ovarian cancer (OR = 0.898, 95% CI: 0.836–0.965, P = 0.003). This association was consistent in direction across supplementary methods (Weighted Median, MR-Egger, Mode-based methods) (Figure 3). A suggestive protective association was observed between Cystatin-C and HGSOC (high-grade serous ovarian carcinoma) risk (OR = 0.908, 95% CI: 0.828–0.994, P = 0.038). In addition, genetically predicted higher levels of Cystatin-F were associated with a lower risk of polycystic ovarian syndrome (PCOS) (OR = 0.852, 95% CI: 0.725–0.999, P = 0.049). No significant causal relationships were identified for any cystatin subtype with ovarian

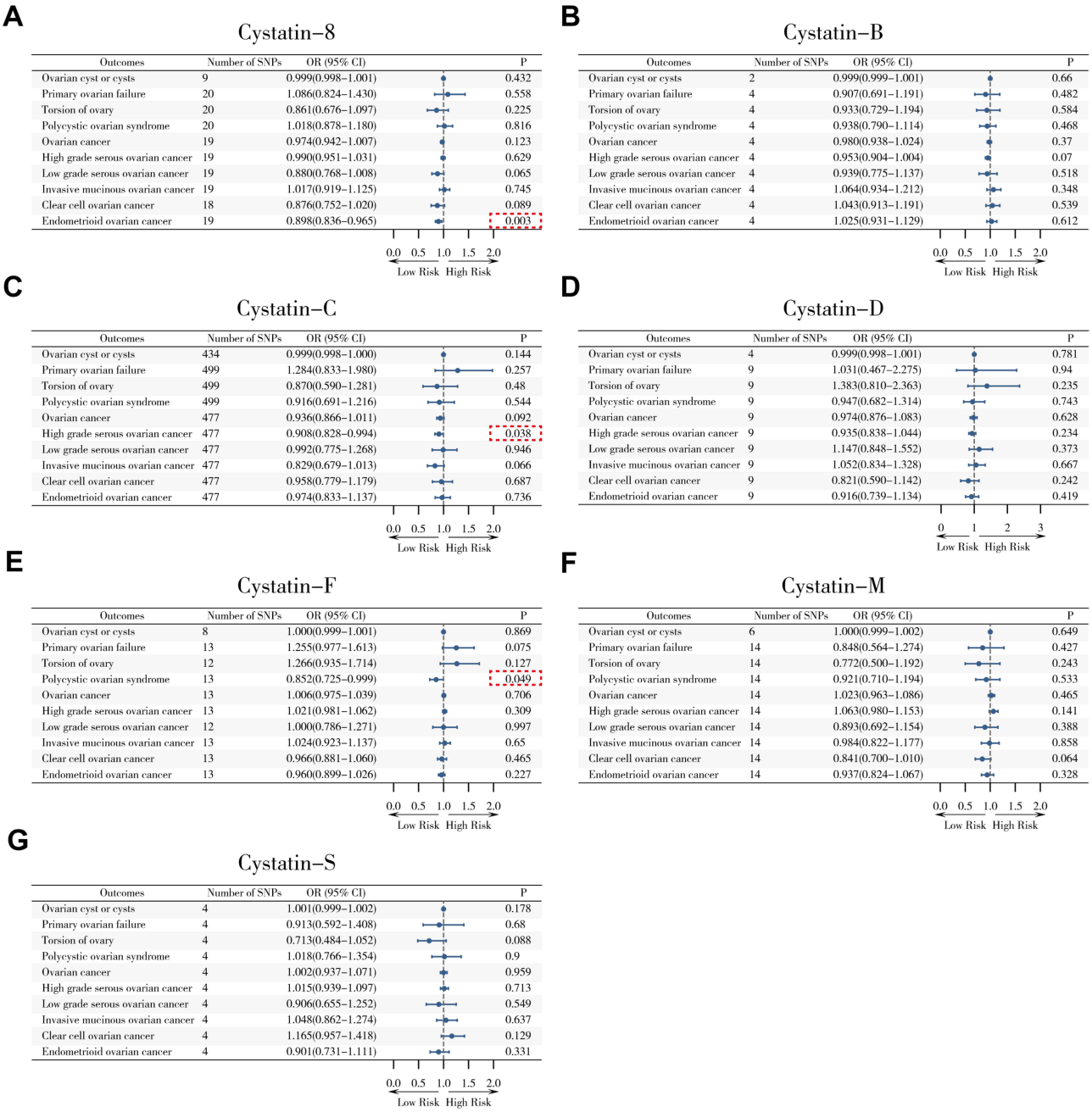

Figure 2 Mendelian randomization estimates for the associations between genetically predicted levels of cystatin subtypes and risks of ovarian lesions. For each cystatin subtype (Cystatin-8 (**A**), Cystatin-B (**B**), Cystatin-C (**C**), Cystatin-D (**D**), Cystatin-F (**E**), Cystatin-M (**F**), Cystatin-S (**G**)), results are presented as odds ratios (OR) with 95% confidence intervals (CIs) per standard deviation increase in genetically predicted cystatin level. The number of SNPs used as instrumental variables is shown for each outcome. Red squares represent significant associations ($P < 0.05$). Horizontal lines represent 95% CIs; squares denote point estimates (OR), with size proportional to the precision of the estimate. A vertical dashed line indicates the null effect (OR = 1). The lower section ("Low Risk/High Risk") provides a reference for interpreting effect direction.

cysts, primary ovarian failure, ovarian torsion, or the other ovarian cancer subtypes (LGSOC, mucinous, clear cell) after multiple testing consideration (all $P_{IVW} \geq 0.05$). Detailed information for all instruments is listed in Figures S1 and S2.

Sensitivity Analyses for the Main MR Estimates

Cochran's Q tests showed no significant heterogeneity for the associations involving Cystatin-8 and endometrioid ovarian cancer (Q = 10.816, $P = 0.902$), as well as Cystatin-F and PCOS (Q = 8.127, $P = 0.775$). Notably, despite significant heterogeneity in the association between Cystatin-C and high-grade serous ovarian cancer as indicated by Cochran's

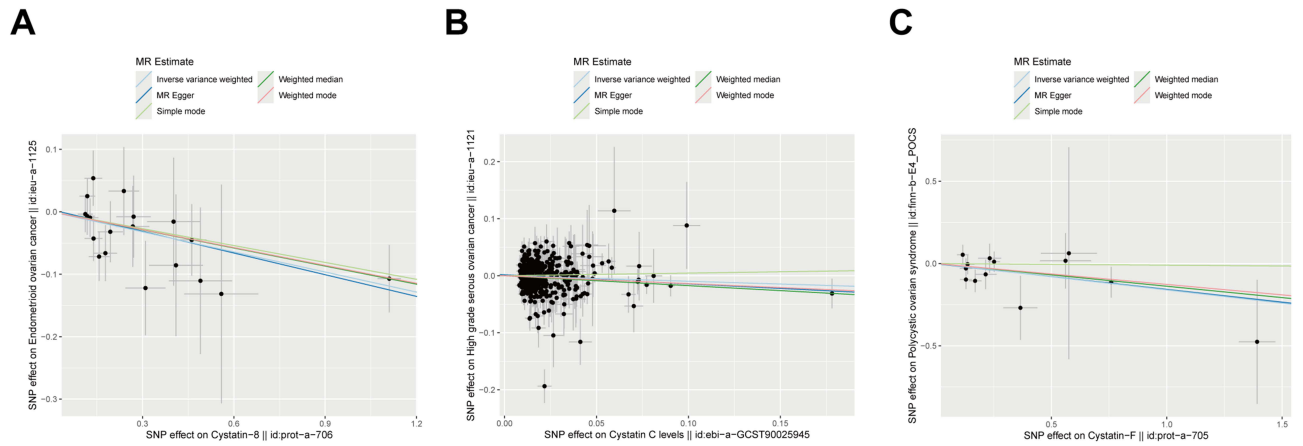

Figure 3 Scatter plot of Mendelian randomization analysis. This scatter plot presents a two-sample MR analysis investigating the potential causal effect of Cystatin-8 on endometrioid ovarian cancer (**A**), Cystatin-C on high-grade serous ovarian carcinoma (**B**) and Cystatin-F on polycystic ovarian syndrome (**C**). Each point represents an SNP. The x-axis shows the SNP-exposure association (genetic effect on Cystatin), and the y-axis shows the SNP-outcome association (genetic effect on ovarian lesions). The regression slope represents the causal estimate per unit change in cystatin levels.

Q test (Q = 626.874, $P < 0.001$), the IVW random-effects model yielded a statistically significant causal estimate (OR = 0.908, 95% CI: 0.828–0.994, $P = 0.038$). The MR-Egger intercept tests yielded intercepts close to zero with non-significant P-values for these associations (all $P_{Egger\ intercept} > 0.1$), indicating no evidence of substantial directional horizontal pleiotropy (Table 2).

To visually assess potential directional pleiotropy and the overall symmetry of the causal estimates, we generated funnel plots for all significant exposure-outcome pairs (eg, Cystatin-8 on ENOC, Cystatin-C on HGSOC and Cystatin-F on PCOS). In a funnel plot, the precision of each instrumental variable (inverse of the standard error) is plotted against its causal estimate (beta coefficient). In the absence of significant pleiotropy or selection bias, the points should be distributed symmetrically around the IVW estimate line. The funnel plots for our primary analyses showed a largely symmetrical distribution of the individual SNP estimates around the pooled IVW estimate line. No pronounced

Table 2 Assessment of the Heterogeneity and Horizontal Pleiotropy Between Cystatin and Ovarian Lesions

Exposures	Outcomes	Heterogeneity		Horizontal Pleiotropy	
		Cochran's Q	*P*	MR-Egger Intercept	*P*
Cystatin-8	Ovarian cyst or cysts	14.601	0.067	0.001	0.213
	Primary ovarian failure	29.796	0.054	−0.029	0.634
	Torsion of ovary	28.229	0.079	0.067	0.202
	Polycystic ovarian syndrome	21.371	0.317	−0.002	0.940
	Ovarian cancer	14.768	0.678	0.004	0.608
	High grade serous ovarian cancer	18.575	0.418	−0.002	0.792
	Low grade serous ovarian cancer	23.650	0.167	0.015	0.616
	Invasive mucinous ovarian cancer	13.180	0.781	−0.003	0.905
	Clear cell ovarian cancer	9.016	0.940	−0.021	0.519
	Endometrioid ovarian cancer	10.816	0.902	0.003	0.820

(*Continued*)

Table 2 (Continued).

Exposures	Outcomes	Heterogeneity		Horizontal Pleiotropy	
		Cochran's Q	*P*	MR-Egger Intercept	*P*
Cystatin-B	Ovarian cyst or cysts	0.268	0.605	NA	NA
	Primary ovarian failure	0.905	0.824	−0.013	0.916
	Torsion of ovary	3.107	0.375	0.041	0.760
	Polycystic ovarian syndrome	1.440	0.696	−0.068	0.426
	Ovarian cancer	1.232	0.745	−0.017	0.430
	High grade serous ovarian cancer	2.381	0.497	−0.023	0.370
	Low grade serous ovarian cancer	4.396	0.222	−0.128	0.171
	Invasive mucinous ovarian cancer	0.702	0.873	0.000	0.996
	Clear cell ovarian cancer	0.284	0.963	0.018	0.756
	Endometrioid ovarian cancer	2.593	0.459	−0.009	0.845
Cystatin-C	Ovarian cyst or cysts	414.705	0.728	2.122e-05	0.252
	Primary ovarian failure	469.767	0.813	−0.008	0.331
	Torsion of ovary	480.789	0.702	0.0129	0.100
	Polycystic ovarian syndrome	524.710	0.197	−0.008	0.143
	Ovarian cancer	633.642	<0.001	−1.361	0.993
	High grade serous ovarian cancer	626.874	<0.001	0.002	0.386
	Low grade serous ovarian cancer	489.888	0.320	−0.010	0.037
	Invasive mucinous ovarian cancer	451.971	0.780	0.002	0.535
	Clear cell ovarian cancer	496.977	0.245	−0.001	0.843
	Endometrioid ovarian cancer	544.302	0.016	−0.003	0.238
Cystatin-D	Ovarian cyst or cysts	3.129	0.372	−0.001	0.532
	Primary ovarian failure	18.593	0.017	−0.245	0.136
	Torsion of ovary	10.850	0.210	−0.026	0.831
	Polycystic ovarian syndrome	3.210	0.920	−0.051	0.469
	Ovarian cancer	14.631	0.067	−0.016	0.461
	High grade serous ovarian cancer	11.084	0.197	−0.001	0.980
	Low grade serous ovarian cancer	9.618	0.293	−0.091	0.133
	Invasive mucinous ovarian cancer	5.258	0.730	−0.020	0.664
	Clear cell ovarian cancer	15.841	0.045	−0.077	0.255
	Endometrioid ovarian cancer	12.231	0.141	0.001	0.978

(*Continued*)

Table 2 (Continued).

Exposures	Outcomes	Heterogeneity		Horizontal Pleiotropy	
		Cochran's Q	*P*	MR-Egger Intercept	*P*
Cystatin-F	Ovarian cyst or cysts	6.430	0.491	6.556e-05	0.573
	Primary ovarian failure	10.068	0.610	0.016	0.782
	Torsion of ovary	9.485	0.577	0.105	0.125
	Polycystic ovarian syndrome	8.127	0.775	−0.002	0.958
	Ovarian cancer	13.010	0.368	−0.003	0.779
	High grade serous ovarian cancer	14.545	0.267	0.003	0.765
	Low grade serous ovarian cancer	24.951	0.009	0.025	0.640
	Invasive mucinous ovarian cancer	14.912	0.246	0.005	0.874
	Clear cell ovarian cancer	5.952	0.918	−0.021	0.426
	Endometrioid ovarian cancer	6.413	0.894	−0.009	0.637
Cystatin-M	Ovarian cyst or cysts	6.218	0.286	−0.002	0.122
	Primary ovarian failure	2.827	0.998	−0.028	0.728
	Torsion of ovary	18.359	0.144	0.036	0.689
	Polycystic ovarian syndrome	7.967	0.846	−0.065	0.225
	Ovarian cancer	8.760	0.791	0.011	0.333
	High grade serous ovarian cancer	16.819	0.208	0.029	0.056
	Low grade serous ovarian cancer	18.281	0.147	−0.050	0.314
	Invasive mucinous ovarian cancer	9.970	0.696	0.029	0.397
	Clear cell ovarian cancer	8.261	0.826	0.014	0.692
	Endometrioid ovarian cancer	8.483	0.811	−0.026	0.313
Cystatin-S	Ovarian cyst or cysts	6.206	0.102	0.001	0.708
	Primary ovarian failure	1.633	0.652	−0.587	0.597
	Torsion of ovary	2.041	0.564	−0.412	0.674
	Polycystic ovarian syndrome	3.195	0.363	−0.089	0.917
	Ovarian cancer	3.169	0.366	0.206	0.284
	High grade serous ovarian cancer	0.171	0.982	0.060	0.757
	Low grade serous ovarian cancer	5.636	0.131	0.839	0.313
	Invasive mucinous ovarian cancer	0.120	0.989	0.081	0.867
	Clear cell ovarian cancer	1.805	0.614	0.264	0.601
	Endometrioid ovarian cancer	6.462	0.091	0.415	0.474

Notes: Cochran's Q test for heterogeneity under the inverse-variance weighted model. MR-Egger intercept test for directional pleiotropy.
Abbreviation: NA, not available.

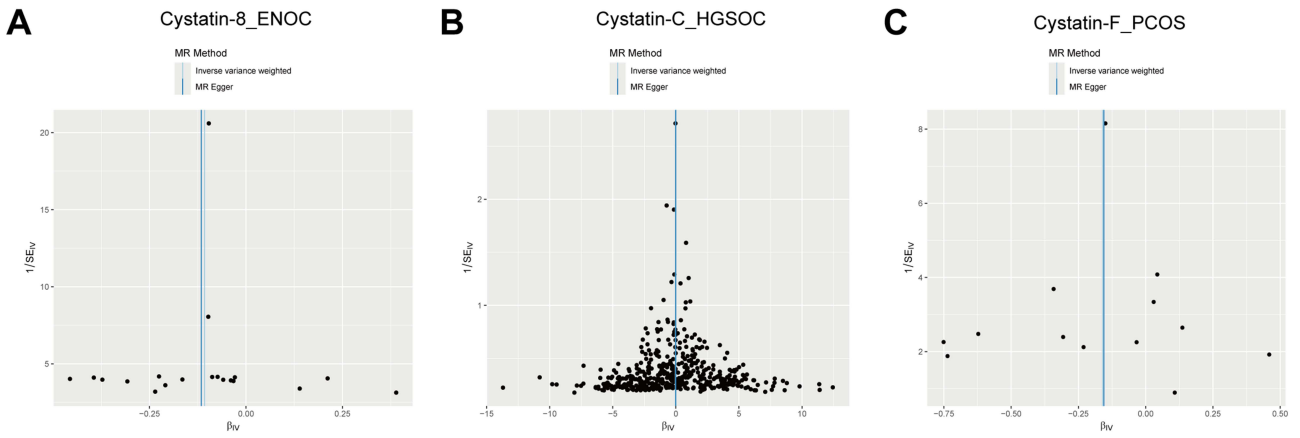

Figure 4 Funnel plot of Mendelian randomization analysis. Funnel plot assessing the potential directional pleiotropy and heterogeneity across SNPs. (**A**) Funnel plot for the causal effect of Cystatin-8 on endometrioid ovarian cancer. (**B**) Funnel plot for the causal effect of Cystatin-C on high-grade serous ovarian cancer. (**C**) Funnel plot for the causal effect of Cystatin-F on polycystic ovary syndrome. Symmetry around the pooled causal estimate (central vertical line) suggests the absence of significant horizontal pleiotropy. Asymmetry may indicate bias or invalid instruments.

asymmetry was observed, which suggests the absence of strong directional pleiotropy that could systematically bias our results (Figure 4). This visual assessment corroborates the non-significant intercepts obtained from the MR-Egger regression, further supporting the validity of the chosen genetic instruments and the robustness of the causal inferences. Besides, the results were not driven by any single influential SNP. The causal estimates remained stable and significant in the same direction after sequentially removing each instrument (Figures S3 and S4). The overall consistency and symmetry observed in both the funnel and scatter plots robustly support the reliability of the causal estimates derived from our Mendelian randomization analysis (Figures S5–S8).

Reverse MR Analysis to Assess Bidirectional Causality

To examine the possibility of reverse causality where the disease outcome might influence cystatin levels, we performed bidirectional MR analyses. For each significant association identified in the forward direction (eg, Cystatin-8 on ENOC, Cystatin-C on HGSOC and Cystatin-F on PCOS), we swapped the exposure and outcome. We selected independent genetic instruments ($P < 5\times10^{-6}$, $R^2 < 0.001$) for the ovarian lesion from the respective GWAS summary statistics and estimated their causal effect on the corresponding cystatin level using the IVW method. A lack of significant association in this reverse direction would support the primary causal inference. The results showed that in all the reverse causal paths tested, no statistically significant causal effects were found. Specifically, the IVW analysis results with endometrioid ovarian cancer as exposure and Cystatin-8 as outcome indicated that the genetically predicted risk of ovarian cancer had no significant effect on Cystatin-8 levels (OR = 0.979, 95% CI: 0.890–1.077, $P = 0.665$). Non-association was observed between HGSOC risk and Cystatin-C (OR = 1.008, 95% CI: 0.957–1.061, $P = 0.777$). For the MR analysis of PCOS on Cystatin-F, a total of eight SNPs were available after harmonization of effect alleles. However, the GWAS summary statistics for the exposure did not report the per-SNP sample size (N), which is a required parameter for calculating the critical *F*-statistic to assess instrument strength. Therefore, we cannot evaluate the risk of weak instrument bias. Consequently, in adherence to methodological best practices for MR, we did not proceed with causal inference for this exposure-outcome pair (Figure 5). Besides, the sensitivity analysis revealed no significant evidence of reverse causality. All estimates and their confidence intervals consistently overlapped the null value, indicating that the observed causal effect of the exposure on the outcome is robust and not confounded by reverse causation (Figure S9).

Discussion

In this comprehensive two-sample MR study, we leveraged large-scale genetic data to investigate the causal roles of seven cystatin subtypes across a spectrum of ovarian lesions. Our primary finding is a robust, genetically predicted protective effect of Cystatin-8 against endometrioid ovarian cancer. We also identified suggestive protective roles for Cystatin-C in high-grade

Outcomes	Exposures	Methods	SNPs (n)	OR (95%CI)	P
ENOC	Cystatin–8	IVW	19	0.898(0.836–0.965)	0.003
		MR Egger	19	0.891(0.804–0.987)	0.041
		Weighted median	19	0.908(0.830–0.993)	0.035
		Simple mode	19	0.914(0.761–1.097)	0.347
		Weighted mode	19	0.909(0.830–0.995)	0.054
HGSOC	Cystatin–C	IVW	477	0.908(0.828–0.994)	0.038
		MR Egger	477	0.853(0.721–1.009)	0.063
		Weighted median	477	0.840(0.724–0.975)	0.022
		Simple mode	477	0.707(0.707–1.552)	0.818
		Weighted mode	477	0.739(0.739–1.027)	0.1
PCOS	Cystatin–F	IVW	13	0.852(0.725–0.999)	0.049
		MR Egger	13	0.856(0.663–1.107)	0.261
		Weighted median	13	0.871(0.707–1.073)	0.195
		Simple mode	13	0.990(0.704–1.392)	0.955
		Weighted mode	13	0.881(0.715–1.084)	0.254
Cystatin–8	ENOC	IVW	10	0.979(0.890–1.077)	0.665
		MR Egger	10	1.067(0.916–1.242)	0.429
		Weighted median	10	1.041(0.915–1.184)	0.542
		Simple mode	10	1.066(0.864–1.315)	0.564
		Weighted mode	10	1.051(0.881–1.254)	0.594
Cystatin–C	HGSOC	IVW	12	1.008(0.957–1.061)	0.777
		MR Egger	12	0.987(0.888–1.098)	0.82
		Weighted median	12	1.002(0.988–1.016)	0.793
		Simple mode	12	0.994(0.975–1.014)	0.569
		Weighted mode	12	1.001(0.987–1.016)	0.889
Cystatin–F	PCOS	IVW			
		MR Egger			
		Weighted median			
		Simple mode			
		Weighted mode			

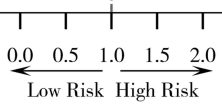

Figure 5 Forest plot of bidirectional Mendelian randomization analysis for cystatin levels on ovarian lesion risk. Each row represents a specific exposure-outcome pair, reporting the number of instrumental SNPs, OR with 95% CI, and *P*-value. No significant associations were observed for endometrioid ovarian cancer or high-grade serous ovarian carcinoma with their respective cystatin exposures. Data for cystatin-F and polycystic ovarian syndrome were unavailable for analysis.

serous ovarian cancer and for Cystatin-F in polycystic ovarian syndrome. These results provide novel genetic evidence highlighting specific members of the cystatin family as potential etiological factors in distinct ovarian pathologies.

Cystatin-8 (CRES, or CST8), a member of the cystatin superfamily, is predominantly and highly expressed in the testis and epididymis functioning in spermatogenesis, sperm maturation, hypothalamic-pituitary-gonadal axis regulation and epididymal antimicrobial defense.[24–26] Current research on Cystatin-8 is largely confined to male reproductive system disorders, with limited and scattered evidence for its low expression in a few malignancies such as bladder cancer,[27] yet no direct original research has reported its expression, functional role or clinical correlation in ovarian lesions, leaving Cystatin-8 a largely unexplored target in ovarian disease. In this study, our findings revealed a significant causal association between Cystatin 8 and endometrioid ovarian cancer risk via the IVW (OR = 0.898, 95% CI: 0.836–0.965, P = 0.003), Weighted Median (OR = 0.908, 95% CI: 0.830–0.993, P = 0.035), and MR Egger (OR = 0.891, 95% CI: 0.804–0.987, P = 0.041) approaches, suggesting that Cystatin 8 may act as a potential protective factor which could pave the way for the development of novel treatments or biomarkers for this cancer subtype. However, further preclinical and clinical studies are needed to elucidate the exact biological mechanisms.

Cystatin-C (CysC, CST3) is a ubiquitous cysteine protease inhibitor that potently suppresses cathepsins, thereby modulating extracellular matrix degradation, inflammation, and cell invasion.[28–31] As a classic and sensitive biomarker for glomerular filtration rate, its role has recently expanded into the realms of tumor biology and metabolic disorders.[32–35] Notably, mounting evidence has clarified the expression pattern, functional role and clinical significance of CysC in ovarian lesions. In ovarian

cancer, CysC is dysregulated in tumor tissues, ascites, serum and cyst fluid, and its abnormal expression is closely linked to tumor progression by regulating the cathepsin-protease inhibitor balance in the tumor microenvironment;[36–38] it is also involved in the pathological processes of polycystic ovary syndrome (PCOS) by mediating insulin resistance and chronic inflammation.[39] These findings confirm CysC as a potential diagnostic or prognostic biomarker and therapeutic target for ovarian diseases, with relevant mechanistic and clinical studies continuously deepening. Our Mendelian randomization analysis identified a significant causal protective effect of CysC on HGSOC based on the IVW method (OR = 0.908, 95% CI: 0.828–0.994, P =0.038) and Weighted Median approach (OR = 0.840, 95% CI: 0.724–0.975, P =0.022), clarifying the genetic causal association between CysC and HGSOC. This finding aligns partially with the observations by Kolwijck et al[40] who reported a strong positive correlation ($P <$ 0.001, R = 0.921) between cathepsin B (CatB) and CysC levels in ovarian cyst fluid (oCF) of malignant serous tumors, and proposed that the increase in CysC might counterbalance CatB-mediated proteolysis, thereby restraining tumor invasiveness. However, in their subsequent study,[36] CysC levels in oCF were not significantly predictive of patient survival. CysC in ovarian cyst fluid likely reflects local biological activity rather than systemic circulation, which may explain the discrepancy between its lack of prognostic significance in oCF and the protective effect observed for circulating CysC in our Mendelian randomization analysis.

The protective effect of Cystatin-F against PCOS is a novel observation (IVW: OR = 0.852, 95% CI: 0.725–0.999, P =0.049). PCOS is a heterogeneous endocrine disorder with chronic low-grade inflammation as a key feature.[41] Cystatin-F is primarily expressed in immune cells and is a potent regulator of cytotoxic lymphocyte activity by inhibiting cathepsin C and other proteases.[42] Higher circulating Cystatin-F might therefore dampen excessive immune activation or modify inflammatory pathways contributing to ovarian dysfunction and metabolic features of PCOS. This finding suggests a previously unrecognized link between immune protease regulation and PCOS pathogenesis.

In contrast, no significant causal relationships were observed between genetically predicted levels of Cystatin B, D, M, or S and any of the examined ovarian outcomes (all $P > 0.05$). This suggests a distinct functional specificity among cystatin family members in ovarian pathophysiology, with only Cystatin-8, -C, and -F demonstrating clear etiological relevance in the studied phenotypes. The isoform-specific effects highlight the importance of precise molecular targeting in future therapeutic strategies.

Our study has several strengths. First, the MR design minimizes residual confounding and reverse causation, providing stronger evidence for causality than observational studies. Second, we analyzed a comprehensive panel of seven cystatin against a wide range of ovarian outcomes, from benign to malignant. Third, the use of large-scale, publicly available GWAS consortia data (OCAC, FinnGen, UK Biobank) ensures substantial statistical power. Fourth, rigorous sensitivity analyses and bidirectional MR assessment confirmed the robustness of our primary findings against pleiotropy and outlier effects. Limitations should also be acknowledged. For example, the genetic instruments were derived from populations of European ancestry, limiting the generalizability of our findings to other ethnic groups. Also, while we used a standard and widely accepted significance threshold ($P < 5\times10^{-6}$), which could marginally increase the risk of including invalid IVs; however, our high F-statistics and consistent sensitivity analyses mitigate this concern. In addition to this, we cannot completely rule out the possibility of undetected horizontal pleiotropy, though the MR-Egger results are reassuring.

Conclusion

In conclusion, this Mendelian randomization study provides novel genetic evidence that higher genetically predicted levels of Cystatin-8 have a causal protective effect against endometrioid ovarian cancer. It also suggests potential protective roles for Cystatin-C in high-grade serous ovarian cancer and for Cystatin-F in polycystic ovarian syndrome. These findings disentangle correlation from causation and highlight specific cystatin as priority candidates for further functional investigation. Future research should focus on elucidating the precise molecular mechanisms by which these protease inhibitors influence ovarian pathophysiology, with the long-term goal of evaluating their potential as biomarkers for risk stratification or as targets for therapeutic intervention in specific ovarian disorders.

Abbreviations

CI, Confidence intervals; EOC, Epithelial ovarian cancer; ENOC, Endometrioid ovarian cancer; HGSOC, High-grade serous ovarian cancer; IVs, Instrumental variables; LGSOC, Low-grade serous ovarian cancer; LD, Linkage

disequilibrium; MOC, Mucinous ovarian cancer; MR, Mendelian randomization; OCCC, Ovarian clear cell cancer; OC, Ovarian cancer; OR, Odds ratios; SNPs, Single nucleotide polymorphisms.

Data Sharing Statement

All data used in this study are publicly available from the MRC-IEU OpenGWAS database, https://opengwas.io/. Analysis scripts are available from the corresponding author upon reasonable request.

Ethics Declaration

The study was submitted to the ethics committee of Fudan University Shanghai Cancer Center. The committee waived the need for ethical approval, as the study used publicly available aggregated data that did not permit re-identification of the original participants. This exemption is supported by Article 32 of the Measures for Ethical Review of Life Science and Medical Research Involving Human Subjects (February 18, 2023, China).

Acknowledgments

We thank all the participants and investigators of the OCAC, FinnGen, and UK Biobank consortia for making their GWAS summary data publicly available.

Funding

This work was supported by the National Natural Science Foundation of China (grant number NSFC-82503569 to T. Y., NSFC-82373383 to R. L.), and the "Young Eagle" Program of the Shanghai Anti-Cancer Association (grant number SACA-CY24C01).

Disclosure

The authors declare no competing interests in this work.

References

1. American College of Obstetricians and Gynecologists. Practice Bulletin No. 174: evaluation and management of adnexal masses. *Obstet Gynecol*. 2016;128(5):e210–e226. doi:10.1097/AOG.0000000000001768
2. Smolarz B, Biernacka K, Łukasiewicz H, et al. Ovarian cancer-epidemiology, classification, pathogenesis, treatment, and estrogen receptors' molecular backgrounds. *Int J Mol Sci*. 2025;26(10):4611. doi:10.3390/ijms26104611
3. Caruso G, Weroha SJ, Cliby W. Ovarian cancer: a review. *JAMA*. 2025;334(14):1278–1291. doi:10.1001/jama.2025.9495
4. Turk V, Stoka V, Vasiljeva O, et al. Cysteine cathepsins: from structure, function and regulation to new frontiers. *Biochim Biophys Acta*. 2012;1824(1):68–88. doi:10.1016/j.bbapap.2011.10.002
5. Mohamed MM, Sloane BF. Cysteine cathepsins: multifunctional enzymes in cancer. *Nat Rev Cancer*. 2006;6(10):764–775. doi:10.1038/nrc1949
6. Jakoš T, Pišlar A, Jewett A, et al. Cysteine cathepsins in tumor-associated immune cells. *Front Immunol*. 2019;10:2037. doi:10.3389/fimmu.2019.02037
7. Wang X, Gui Lu, Zhang Y, et al. Cystatin B is a progression marker of human epithelial ovarian tumors mediated by the TGF-β signaling pathway. *Int J Oncol*. 2014;44(4):1099–1106. doi:10.3892/ijo.2014.2261
8. Levin MG, Burgess S. Mendelian randomization as a tool for cardiovascular research: a review. *JAMA Cardiol*. 2024;9(1):79–89. doi:10.1001/jamacardio.2023.4115
9. Emdin CA, Khera AV, Kathiresan S. Mendelian randomization. *JAMA*. 2017;318(19):1925–1926. doi:10.1001/jama.2017.17219
10. Marchal MI, Mutie PM, Azimi T, et al. Distinct genetic profiles of obesity have different effects on breast cancer risk: leveraging Mendelian randomisation to interrogate causal pathways and identify mediating proteins. *Breast Cancer Res*. 2026;28(1):23. doi:10.1186/s13058-025-02214-3
11. Li ZY, Li H-L, Wang J, et al. Metabolomic insights into associations between adiposity markers and liver cancer risk: results from a prospective cohort study and Mendelian randomization analysis. *PLoS Med*. 2026;23(2):e1004910. doi:10.1371/journal.pmed.1004910
12. Dong H, Nie S, Ni N, et al. Exploring the potential causal relationship of galectin levels and gynecologic cancers: a two-sample Mendelian randomization study. *Int J Womens Health*. 2025;17:5017–5034. doi:10.2147/IJWH.S515826
13. Ying J, Chen X, Lv T, et al. Mendelian randomization analysis to explore the relationship between cathepsins and malignant ovarian tumors. *Medicine*. 2024;103(46):e40219. doi:10.1097/MD.0000000000040219
14. Skrivankova VW, Richmond RC, Woolf BAR, et al. Strengthening the reporting of observational studies in epidemiology using Mendelian randomization: the STROBE-MR statement. *JAMA*. 2021;326(16):1614–1621. doi:10.1001/jama.2021.18236
15. Folkersen L, Fauman E, Sabater-Lleal M, et al. Mapping of 79 loci for 83 plasma protein biomarkers in cardiovascular disease. *PLoS Genet*. 2017;13(4):e1006706. doi:10.1371/journal.pgen.1006706
16. Barton AR, Sherman MA, Mukamel RE, et al. Whole-exome imputation within UK Biobank powers rare coding variant association and fine-mapping analyses. *Nat Genet*. 2021;53(8):1260–1269. doi:10.1038/s41588-021-00892-1
17. Suhre K, Arnold M, Bhagwat AM, et al. Connecting genetic risk to disease end points through the human blood plasma proteome. *Nat Commun*. 2017;8(1):14357. doi:10.1038/ncomms14357
18. Sun BB, Maranville JC, Peters JE, et al. Genomic atlas of the human plasma proteome. *Nature*. 2018;558(7708):73–79. doi:10.1038/s41586-018-0175-2

19. 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(5):680–691. doi:10.1038/ng.3826
20. Bowden J, Del Greco M F, Minelli C, et al. Assessing the suitability of summary data for two-sample Mendelian randomization analyses using MR-Egger regression: the role of the I2 statistic. *Int J Epidemiol*. 2016;45(6):1961–1974. doi:10.1093/ije/dyw220
21. Zuber V, Colijn JM, Klaver C, et al. Selecting likely causal risk factors from high-throughput experiments using multivariable Mendelian randomization. *Nat Commun*. 2020;11(1):29. doi:10.1038/s41467-019-13870-3
22. Zheng J, Baird D, Borges M-C, et al. Recent developments in Mendelian randomization studies. *Curr Epidemiol Rep*. 2017;4(4):330–345. doi:10.1007/s40471-017-0128-6
23. Bowden J, Davey Smith G, Haycock PC, et al. Consistent estimation in Mendelian randomization with some invalid instruments using a weighted median estimator. *Genet Epidemiol*. 2016;40(4):304–314. doi:10.1002/gepi.21965
24. Ferrer M, Cornwall G, Oko R. A population of CRES resides in the outer dense fibers of spermatozoa. *Biol Reprod*. 2013;88(3):65. doi:10.1095/biolreprod.112.104745
25. Sutton HG, Fusco A, Cornwall GA. Cystatin-related epididymal spermatogenic protein colocalizes with luteinizing hormone-beta protein in mouse anterior pituitary gonadotropes. *Endocrinology*. 1999;140(6):2721–2732. doi:10.1210/endo.140.6.6777
26. Wang L, Yuan Q, Chen S, et al. Antimicrobial activity and molecular mechanism of the CRES protein. *PLoS One*. 2012;7(11):e48368. doi:10.1371/journal.pone.0048368
27. Xu H, Liu W, Kuang X, et al. Expression profiles and clinical significance of cystatin family genes in transitional cell carcinoma of the urinary bladder. *Bladder*. 2025;12(1):e21200035. doi:10.14440/bladder.2024.0057
28. Xu Y, Ding Y, Li X, et al. Cystatin C is a disease-associated protein subject to multiple regulation. *Immunol Cell Biol*. 2015;93(5):442–451. doi:10.1038/icb.2014.121
29. Xie L, Terrand J, Xu B, et al. Cystatin C increases in cardiac injury: a role in extracellular matrix protein modulation. *Cardiovasc Res*. 2010;87(4):628–635. doi:10.1093/cvr/cvq138
30. Dedual MA, Wueest S, Challa TD, et al. Obesity-induced increase in cystatin C alleviates tissue inflammation. *Diabetes*. 2020;69(9):1927–1935. doi:10.2337/db19-1206
31. Carlsson E, Supharattanasitthi W, Jackson M, et al. Increased rate of retinal pigment epithelial cell migration and pro-angiogenic potential ensuing from reduced cystatin C expression. *Invest Ophthalmol Vis Sci*. 2020;61(2):9. doi:10.1167/iovs.61.2.9
32. Sun H, Wu Z, Wang G, et al. Normalized creatinine-to-cystatin C ratio and risk of cardiometabolic multimorbidity in middle-aged and older adults: insights from the China health and retirement longitudinal study. *Diabetes Metab J*. 2025;49(3):448–461. doi:10.4093/dmj.2024.0100
33. Song X, Xiong L, Guo T, et al. Cystatin C is a predictor for long-term, all-cause, and cardiovascular mortality in US adults with metabolic syndrome. *J Clin Endocrinol Metab*. 2024;109(11):2905–2919. doi:10.1210/clinem/dgae225
34. Xiong K, Zhang S, Zhong P, et al. Serum cystatin C for risk stratification of prediabetes and diabetes populations. *Diabetes Metab Syndr*. 2023;17(11):102882. doi:10.1016/j.dsx.2023.102882
35. Jung CY, Kim HW, Han SH, et al. Creatinine-cystatin C ratio and mortality in cancer patients: a retrospective cohort study. *J Cachexia Sarcopenia Muscle*. 2022;13(4):2064–2072. doi:10.1002/jcsm.13006
36. Kolwijck E, Kos J, Obermajer N, et al. The balance between extracellular cathepsins and cystatin C is of importance for ovarian cancer. *Eur J Clin Invest*. 2010;40(7):591–599. doi:10.1111/j.1365-2362.2010.02305.x
37. Bodnar L, Wcislo GB, Smoter M, et al. Cystatin C as a parameter of glomerular filtration rate in patients with ovarian cancer. *Kidney Blood Press Res*. 2010;33(5):360–367. doi:10.1159/000319097
38. Nishikawa H, Ozaki Y, Nakanishi T, et al. The role of cathepsin B and cystatin C in the mechanisms of invasion by ovarian cancer. *Gynecol Oncol*. 2004;92(3):881–886. doi:10.1016/j.ygyno.2003.11.017
39. Özdemir Başer Ö, Göçmen AY, Aydoğan Kırmızı D. The role of inflammation, oxidation and Cystatin-C in the pathophysiology of polycystic ovary syndrome. *Turk J Obstet Gynecol*. 2022;19(3):229–235. doi:10.4274/tjod.galenos.2022.29498
40. Kolwijck E, Massuger LFAG, Thomas CMG, et al. Cathepsins B, L and cystatin C in cyst fluid of ovarian tumors. *J Cancer Res Clin Oncol*. 2010;136(5):771–778. doi:10.1007/s00432-009-0716-z
41. Escobar-Morreale HF. Polycystic ovary syndrome: definition, aetiology, diagnosis and treatment. *Nat Rev Endocrinol*. 2018;14(5):270–284. doi:10.1038/nrendo.2018.24
42. Hamilton G, Colbert JD, Schuettelkopf AW, et al. Cystatin F is a cathepsin C-directed protease inhibitor regulated by proteolysis. *EMBO J*. 2008;27(3):499–508. doi:10.1038/sj.emboj.7601979