Considine, D. P. C., Jia, G., Shu, X., Pharoah, P. D. P., Zheng, W., & Kar, S. (2020). Genetically predicted circulating protein biomarkers and ovarian cancer risk. *Gynecologic Oncology*. https://doi.org/10.1016/j.ygyno.2020.11.016

Genetically predicted circulating protein biomarkers and ovarian cancer risk

Daniel P. C. Considine*[1], Guochong Jia*[2], Xiang Shu[2,3], Joellen M. Schildkraut[4], Paul D. P. Pharoah†[1,5], Wei Zheng†[2], and Siddhartha P. Kar†++[6,7] for the Ovarian Cancer Association Consortium

1. Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK.
2. Vanderbilt Epidemiology Center, Division of Epidemiology, Department of Medicine, Vanderbilt-Ingram Cancer Center, Vanderbilt University Medical Center, Nashville, TN, USA.
3. Department of Epidemiology & Biostatistics, Memorial Sloan Kettering Cancer Center, New York, NY, USA.
4. Department of Epidemiology, Rollins School of Public Health, Emory University, Atlanta, GA, USA.
5. Centre for Cancer Genetic Epidemiology, Department of Oncology, University of Cambridge, Cambridge, UK.
6. MRC Integrative Epidemiology Unit, University of Bristol, Bristol, UK.
7. Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK.

*These authors contributed equally and are co-first authors.
†These authors contributed equally and are co-last authors.
++Indicates corresponding author.

For correspondence:

Siddhartha P. Kar, MBBS, PhD, MPH
MRC Integrative Epidemiology Unit
BS11, Oakfield House, Oakfield Grove
University of Bristol, Bristol BS8 2BN, UK
Email: siddhartha.kar@bristol.ac.uk

Word counts: Abstract = 252, Manuscript = 3,646

Abstract

Objective

Most women with epithelial ovarian cancer (EOC) are diagnosed after the disease has metastasized and survival in this group remains poor. Circulating proteins associated with the risk of developing EOC have the potential to serve as biomarkers for early detection and diagnosis. We integrated large-scale genomic and proteomic data to identify novel plasma proteins associated with EOC risk.

Methods

We used the germline genetic variants most strongly associated ($P<1.5\times10^{-11}$) with plasma levels of 1,329 proteins in 3,301 healthy individuals from the INTERVAL study to predict circulating levels of these proteins in 22,406 EOC cases and 40,941 controls from the Ovarian Cancer Association Consortium (OCAC). Association testing was performed by weighting the beta coefficients and standard errors for EOC risk from the OCAC study by the inverse of the beta coefficients from INTERVAL.

Results

We identified 26 proteins whose genetically predicted circulating levels were associated with EOC risk at false discovery rate<0.05. The 26 proteins included MFAP2, SEMG2, DLK1, and NTNG1 and a group of 22 proteins whose plasma levels were predicted by variants at chromosome 9q34.2. All 26 protein association signals identified were driven by association with the high-grade serous histotype that comprised 58% of the EOC cases in OCAC. Regional genomic plots confirmed overlap of the genetic association signal underlying both plasma protein level and EOC risk for the 26 proteins. Pathway analysis identified enrichment of seven biological pathways among the 26 proteins ($P_{adjusted}<0.05$), highlighting roles for Focal Adhesion-PI3K-Akt-mTOR and Notch signaling.

Conclusion

The identified proteins further illuminate the etiology of EOC and represent promising new EOC biomarkers for targeted validation by studies involving direct measurement of plasma proteins in EOC patient cohorts.

Keywords

epithelial ovarian cancer, risk, circulating proteins, circulating biomarkers, genome-wide association study

Research Highlights

- This study analyzed 667 germline genetic variants known to be associated with circulating (plasma) levels of 1,329 proteins
- These variants were used to predict plasma protein levels in 22,406 epithelial ovarian cancer cases and 40,941 controls
- Genetically predicted levels of 26 proteins were associated with all invasive epithelial ovarian cancer risk
- The identified proteins were enriched for the Focal Adhesion-PI3K-Akt-mTOR signaling, Notch signaling, and other pathways
- The identified proteins have the potential to serve as circulating biomarkers particularly for high-grade serous epithelial ovarian cancer risk

Introduction

Ovarian cancer is the most common cause of death from gynecological malignancy in the United States and accounted for an estimated 295,000 incident cases and 184,000 deaths globally in 2018 [1,2]. Despite advances in treatment, survival rates in ovarian cancer continue to remain low, in part, due to the late detection of most cases [3]. Nearly four decades after its discovery [4], circulating levels of the protein cancer antigen 125 (CA-125) continue to be used to screen women at high risk of developing ovarian cancer, such as those with a hereditary cancer syndrome, and women with abnormal findings on examination and/or ultrasound. However, CA-125 has limited sensitivity and specificity in these settings [5]. Furthermore, screening asymptomatic women for CA-125 level, despite the use of serial measurements and algorithmic approaches to the interpretation of these levels [6] – and even in combination with transvaginal ultrasound – does not reduce ovarian cancer mortality and is not recommended by the US Preventive Services Task Force [7]. Human epididymis secretory protein E4 (HE4) has been developed in recent years as a blood-based protein biomarker for the diagnosis of ovarian carcinoma [8], and the combination of CA-125 and HE4 is a more accurate predictor of ovarian malignancy than either biomarker alone [9]. However, there remains an urgent unmet need to identify novel circulating protein biomarkers that will be more useful for the early detection of this aggressive disease.

Studies in search of new plasma protein biomarkers in ovarian cancer have been restricted to small sample sizes and evaluated limited protein panels [10,11]. In the current study, we adopted a different approach to the identification of circulating protein biomarkers of ovarian cancer risk using large-scale data from two genome-wide association studies (GWAS). The first data set was a GWAS of healthy blood donors in the INTERVAL study that has identified robust associations between inherited genetic variants and plasma protein levels [12]. The second data set was the the largest and latest published GWAS meta-analysis from the Ovarian Cancer Association Consortium

(OCAC) [13]. While epithelial ovarian cancer (EOC) accounts for approximately 90% of all ovarian cancer cases, EOC itself is a diverse entity with distinct histological subtypes: high-grade serous (the most common and lethal histotype), low-grade serous, clear cell, mucinous, endometrioid, and low malignant potential (serous or mucinous) tumors. The OCAC GWAS included associations with all invasive and histotype-specific EOC susceptibility. We used the inherited genetic variants robustly associated with plasma protein levels in the INTERVAL GWAS to predict these levels in the OCAC GWAS where plasma protein levels have not actually been measured but the variants have been genotyped. Such predictions are likely to suffer from less selection bias and confounding because the genetic variants on which they are based are randomly allocated at gametogenesis and fixed after conception. Our study design enabled a comprehensive appraisal of the role of the levels of over 1,300 plasma proteins in more than 22,000 ovarian cancer cases and over 40,000 controls.

Methods

Circulating (plasma) protein data set

We used effect size estimates (beta coefficients) from genome-wide association analyses linking 667 single nucleotide polymorphisms (SNPs) to the circulating (plasma) levels of 1,329 proteins in 3,301 healthy participants from the INTERVAL study [12], a bioresource of blood donors in England who were recruited into a multi-center randomized trial of blood donation frequency [14]. Each of these SNPs was associated with at least one of the plasma proteins at genome-wide significance (defined as $P < 1.5 \times 10^{-11}$ in the INTERVAL analysis [12]) and was the SNP most strongly associated with the circulating levels of that protein. Five hundred and eight-five SNPs were associated with the levels of only one circulating protein each while 82 SNPs were associated with multiple proteins (ranging from 36 SNPs that were associated with two proteins each to one SNP that was associated with 95 proteins; Table S1). We restricted analysis to SNPs that had minor allele frequency (MAF) > 1% and

had either been genotyped or imputed with quality score > 0.8 – both in the INTERVAL analytic sample and in the OCAC data set (described below). These SNP-protein associations included 908 *trans*-associations where the top SNP associated with the protein was > 1 Mb away from the gene encoding the protein and 421 *cis*-associations where the top SNP associated with the protein was < 1 Mb away from the gene encoding the protein. Plasma protein levels in the INTERVAL study were quantified using an expanded aptamer-based multiplex protein assay called SOMAscan [12,15] and germline genotypes were measured on Affymetrix Axiom UK Biobank array with imputation into a combined combined 1000 Genomes Phase 3-UK10K reference panel. We used the same protein names and identifiers, including UniProt and SOMAmer IDs (Table S1), as used in the original INTERVAL genetic report [12] for consistency. That report contains additional details of sample and genotype quality control, imputation, and association analysis in the INTERVAL study.

Epithelial ovarian cancer data set

Summary statistics (beta coefficients and standard errors) from a GWAS meta-analysis for EOC susceptibility in women of European ancestry were obtained from OCAC [13]. The GWAS meta-analysis included 22,406 invasive EOC cases overall and 40,941 controls and this “all invasive EOC” case-control set was the focus of the primary analysis in the current study. EOC histotype-specific summary statistics from the same GWAS meta-analysis were also evaluated for seven histological subtypes as a secondary analysis in the current study. This included high-grade serous (13,037 cases), low-grade serous (1,012 cases), low malignant potential serous (1,954 cases), invasive mucinous (1,417 cases), low malignant potential mucinous (1,149), clear cell (1,366), and endometrioid (2,810 cases) EOC cases and 40,941 controls. Additional details of sample and genotype quality control, imputation, and association analytic procedures for the OCAC GWAS meta-analysis have been previously published [13].

Statistical analysis

We used the Wald ratio to estimate the effect of genetically predicted circulating protein levels on ovarian cancer risk. The Wald estimator in this context is the ratio of the beta coefficient for a SNP from the ovarian cancer GWAS meta-analysis to the beta coefficient for the same SNP from the plasma protein genome-wide association analysis. The SNP most strongly associated with the circulating level of each protein in the INTERVAL data set was used. The standard error of the Wald estimator is the ratio of the standard error for the SNP from the ovarian cancer GWAS meta-analysis to the absolute value of the beta coefficient for the SNP from the plasma protein genome-wide association analysis. These analyses were performed using the *R* (version 3.6.2) statistical computing language. *P*-values were calculated using the formula: "pnorm(abs(Wald_estimator)/standard_error_of_Wald_estimator, lower.tail=FALSE) * 2" and the multiple comparisons burden for testing 1,329 SNP-protein-ovarian cancer associations was accounted for using false discovery rate (FDR) control by the method of Benjamini and Hochberg as implemented in the "p.adjust" function. The Wald estimator allowed for incorporation of the beta coefficient for the SNP from the plasma protein analysis and allowed easy inference of the direction of the association (whether positive or inverse) between plasma protein level and ovarian cancer risk. Therefore, we preferred the Wald estimator over directly testing for the genetic association between the top plasma protein level-associated SNP and ovarian cancer risk (although in practice both approaches provided almost identical *P*-values). As noted above, our primary analysis was for all invasive EOC risk, given that this combined phenotype had the largest sample size, while in secondary analyses we evaluated histotype-specific risk.

We followed up genetically predicted circulating levels of proteins that were found to be associated (FDR < 0.05) in our study with ovarian cancer risk to assess whether the top plasma protein level-associated SNP was part of the top ovarian cancer genetic association signal in the

same genomic region – a positional overlap that would reinforce the role of the SNP as a driver of both circulating protein levels and ovarian cancer risk. We did this by visualizing ovarian cancer genetic associations for all SNPs with MAF > 1% and imputation quality > 0.8 in the OCAC data set in the 500 kb window centered on the top protein-associated SNP (i.e., +/- 250 kb on either side) using two-way scatter plots generated in *Stata* (version 14, StataCorp LP, College Station, TX). For SNPs with stronger *P*-values for association with ovarian cancer risk in OCAC as compared to the top protein-associated SNP, the correlation between the stronger *P*-value SNPs and the top protein-associated SNP was calculated using the LDlink online tool and data from the 1000 Genomes European ancestry populations [16]. If the same SNP association signal drives both plasma protein level and ovarian cancer risk, we expected one of the following three scenarios to be true: (i) the top protein-associated SNP is also the top ovarian cancer associated SNP or (ii) it is strongly correlated ($r^2 > 0.9$) with the top ovarian cancer associated SNP(s) or (iii) there are multiple independent ($r^2 < 0.01$) genetic association signals in OCAC in the same region and the top protein-associated SNP is one of these associations. A second follow-up analysis of proteins that achieved FDR < 0.05 in our study involved mapping these to the genes encoding them and evaluating the genes for enrichment of pathways (at $P < 0.05$ after adjustment for testing multiple pathways) using the Enrichr online tool [17] and the "WikiPathways Human 2019" database [18] that contains annotations for 472 known biological pathways. A final follow-up analysis involved searching for genome-wide signficant associations ($P < 5 \times 10^{-8}$) between the top plasma protein level-associated SNP for each of the proteins that achieved FDR < 0.05 in our study and other diseases and traits in the published (i.e., MEDLINE indexed) literature. This search was performed using the PhenoScanner (version 2) online tool [19], querying published European-ancestry GWAS. The aim was to identify pleiotropic diseases and traits that may provide an alternative explanation for the plasma protein-EOC risk associations identified, stemming from their associations with the same top SNPs. Such pleiotropic diseases and traits associated with the same SNPs may also be the cause or consequence of plasma protein level changes that in turn are associated with EOC risk.

Results

Genetically predicted circulating levels of 26 proteins were associated with all invasive EOC risk at FDR < 0.05 (13 positive and 13 inverse associations; Table 1 and Table S1). First, this included a positive association between MFAP2 encoded by *MFAP2* on chromosome 1 and all invasive EOC risk (Microfibrillar-associated protein 2; $P_{Wald} = 1.8 \times 10^{-4}$, FDR = 0.01). The top MFAP2 plasma protein-associated SNP rs4920605 ($P_{OCAC\text{-}GWAS\text{-}all} = 1.82 \times 10^{-4}$) in the INTERVAL study was ~8 kb from the transcription start site (TSS) of *MFAP2*. There was only one SNP in the same region, rs143483351 ($P_{OCAC\text{-}GWAS\text{-}all} = 1.76 \times 10^{-4}$), a multi-allelic variant 2 kb from rs4920605, with a slightly stronger association with all invasive EOC risk (Fig. 1 (a) and Table S2). SNP rs143483351 could not be evaluated in LDlink [16] for correlation with rs4920605 because it was a multi-allelic variant. Second, our FDR < 0.05 results also included an inverse association between NTNG1 encoded by *NTNG1* on chromosome 1 (in a genomic region distinct from *MFAP2*) and all invasive EOC risk (Netrin-G1; $P_{Wald} = 4.9 \times 10^{-4}$, FDR = 0.03). The top NTNG1 plasma protein-associated SNP rs115668827 ($P_{OCAC\text{-}GWAS\text{-}all} = 4.9 \times 10^{-4}$) in the INTERVAL study was ~4 kb from the TSS of *NTNG1*. There was only one SNP in the same region, rs11185086 ($P_{OCAC\text{-}GWAS\text{-}all} = 3.8 \times 10^{-4}$), 173 kb from rs115668827, with a stronger association with all invasive EOC risk (Fig. 1 (b) and Table S2). However, rs11185086 and rs115668827 represented independent signals in the same region ($r^2 = 7 \times 10^{-4}$). Third, the list of 26 plasma proteins identified included positive associations between SEMG2 (Semenogelin-2) and ovarian cancer risk and DLK1 (Protein delta homolog 1) and ovarian cancer risk (for both associations – $P_{Wald} = 4.0 \times 10^{-4}$, FDR = 0.02). The top SEMG2 plasma protein-associated SNP and the top DLK1 plasma protein-associated SNP in the INTERVAL study was the same SNP, rs12881760 ($P_{OCAC\text{-}GWAS\text{-}all} = 3.96 \times 10^{-4}$), which is ~16 kb from the TSS of *DLK1* on chromosome 14. SEMG2 is encoded by *SEMG2* on chromosome 20 and rs12881760 is associated with its circulating level by acting in *trans*. SNP rs12881760 is part of a cluster of three SNPs that includes rs10144381

($P_{OCAC\text{-}GWAS\text{-}all}$ = 3.50 x 10^{-4}) and rs12881545 ($P_{OCAC\text{-}GWAS\text{-}all}$ = 3.56 x 10^{-4}), which are within 3 kb of each other and strongly correlated (r^2 > 0.93), and together mark the strongest association signal with all invasive EOC risk in the *DLK1* region (Fig. 1 (c) and Table S2).

The remaining 22 of the 26 all invasive EOC risk-associated circulating proteins identified were proxied by 10 correlated SNPs (r^2 > 0.38) spanning a ~10 kb interval on chromosome 9 (Table 1). Three of these SNPs were the top SNP for one protein each, four for two proteins each, two for three proteins each, and one SNP was the top SNP for five proteins in the INTERVAL data set (Table 1). Ten proteins demonstrated a positive association and 12 showed an inverse association with all invasive EOC risk. Twenty-one of the 22 proteins were encoded by genes > 1 Mb away from this chromosome 9 interval (*trans*-associations) and most were in fact encoded by genes located on other chromosomes. The only exception to this was the plasma protein BGAT (Histo-blood group ABO system transferase) encoded by *ABO* and the top BGAT plasma level-associated SNP, rs505922, is ~2 kb from the TSS of *ABO*. The *ABO* locus (chromosome 9q34.2) is a known genome-wide significant ($P < 5 \times 10^{-8}$) locus for all invasive and high-grade serous ovarian cancer risk [13,20]. The ten protein level-associated SNPs spanned the *ABO* locus and were among the top 50 all invasive EOC risk SNPs in the 500 kb region (Table S2 and Fig. 1 (d)). The top all invasive EOC risk SNP in the region, rs587729126 ($P_{OCAC\text{-}GWAS\text{-}all}$ = 8.3 x 10^{-10}), was the top SNP in the INTERVAL study for association with circulating levels of FA20B (Glycosaminoglycan xylosylkinase) and sICAM-2 (Intercellular adhesion molecule 2). This overlap of top associations led to these two proteins emerging as the plasma proteins whose genetically predicted levels were most strongly associated with all invasive EOC risk in our analysis (for both associations – P_{Wald} = 8.1 x 10^{-10}, FDR = 4.5 x 10^{-7}; Table 1). The PhenoScanner search indicated that eight of the ten protein level-associated SNPs that spanned the *ABO* locus were associated with 62 traits (Table S3). Overall, for all 26 proteins identified (associated with SNPs in the regions presented in Fig. 1 and discussed above), we observed a clear overlap between the top circulating protein level-associated SNP and the top all invasive EOC risk association, lending further confidence to the

association between plasma protein levels predicted by these SNPs and disease risk. We did not identify any additional proteins at FDR < 0.05 in any of the histotype-specific analyses (Table S1). An inspection of the high-grade serous EOC results (Table S1) confirmed that all 26 FDR < 0.05 protein associations with all invasive EOC risk were driven by associations in the high-grade serous EOC sample, which contributed the largest number of cases to the all invasive EOC sample. Pathway enrichment analysis of the genes encoding the 26 proteins identified seven pathways at $P_{adjusted} < 0.05$ (Table 2).

Discussion

By combining genome-wide association data from 22,406 all invasive EOC cases and 40,941 controls and plasma proteome-wide genetic association data from 3,301 healthy individuals, we identified 26 proteins whose genetically inferred circulating levels were associated with EOC risk after false discovery rate control (FDR < 0.05). The combination of these data sets offered unprecedented scale to evaluate the role of over 1,300 plasma proteins in the development of EOC and identified circulating protein biomarkers with the potential for clinical translational in the early detection and diagnosis of EOC.

We observed that the top plasma protein level-associated SNP was either the top all invasive EOC risk SNP in the 500 kb region centred on the SNP (Fig. 1 (a) and (d)) or it was the top SNP of one of two independent ($r^2 < 0.01$) all invasive EOC risk associations in the region (Fig. 1 (b)) or it was part of a cluster of highly correlated ($r^2 > 0.9$) SNPs that together marked the top all invasive EOC risk association in the region (Fig. 1 (c)). This suggests that our results are unlikely to be due to linkage disequilibrium contamination, i.e., regional genetic architecture where the top plasma protein level-associated SNP is weakly correlated with the top all invasive EOC risk SNP and results in a spurious association underpinned by two distinct SNP signals (one for protein and another for EOC). While

363 the focus of our analysis was the use of SNPs associated with plasma protein levels to evaluate the
364 association between plasma protein levels and EOC risk and not the direct genetic association
365 between SNPs and EOC risk, we note that there were 667 unique SNPs used in the analysis (Table S1)
366 and the 13 unique SNPs underpinning the 26 proteins identified (Table 1) were all associated with all
367 invasive EOC risk at $P < 0.05/667$, which would be the conventional threshold for statistical
368 significance if this was a SNP-based association study of 667 SNPs. Ten of the 13 SNPs were
369 genome-wide significant ($P < 5 \times 10^{-8}$) as they are located at a previously reported all invasive and
370 high-grade serous EOC risk locus at or near *ABO* on chromosome 9q34.2 [13,20]. The three
371 remaining SNPs (spanning three distinct genomic regions; Fig. 1 (a), (b), and (c)) may well represent
372 as yet unidentified genetic susceptibility loci for all invasive EOC that we are presently
373 underpowered to detect at GWAS levels of significance ($P < 5 \times 10^{-8}$). Thus, loci known to be
374 associated with the plasma proteome may aid in the discovery of sub-threshold GWAS loci for
375 disease susceptibility in much the same way as previously demonstrated for other biological
376 information integrated into genetic association studies [21].
377
378 Pathway analysis highlighted that five of the 26 proteins whose genetically predicted plasma
379 levels were associated with all invasive EOC risk at FDR < 0.05 belonged to the "Focal Adhesion-PI3K-
380 Akt-mTOR-signaling pathway ($P_{pathway\ (adjusted)} = 0.006$; Table 2), which was the maximum overlap seen
381 between any established biomolecular pathway and the 26 proteins. The genes encoding these
382 proteins were located across different chromosomes, but the SNPs associated most strongly with
383 their plasma levels were all located at the 9q34.2 locus. The PI3K-Akt-mTOR intracellular signaling
384 cascade is a major regulator of the cell cycle and has key roles in cellular quiescence, growth and
385 proliferation, and cancer cell survival and metastasis [22]. Somatic aberrations in this pathway are
386 found in the majority of high-grade serous ovarian tumors [23]. Another pathway identified at
387 $P_{pathway\ (adjusted)} < 0.05$ was Notch signaling and this association was driven, in turn, by associations
388 between genetically predicted circulating levels of two Notch proteins, MFAP2 (chromosome

1p36.13) and DLK1 (chromosome 14q32.2), and all invasive EOC risk. It is noteworthy that these two plasma proteins associated with EOC risk at FDR < 0.05 were encoded by genes located on distinct chromosomes but the genes/proteins were members of the same biological pathway. DLK1, a non-canonical Notch ligand, has a demonstrated role in promoting ovarian carcinogenesis via Notch activation and epithelial-mesenchymal transition [24]. The microfibrillar-associated protein 2 (MFAP2), previously named microfibril-associated glycoprotein 1 (MAGP1), activates integrin signaling and is a potential oncogene [25–27]. Another protein identified at FDR < 0.05, Netrin-G1 or NTNG1, is involved in apoptosis and known to be dysregulated particularly in endocrine-related tumors [28,29]. Further, the gene that encodes NTNG1 has been shown to be overexpressed in malignant ovarian tumors [30].

Larger genetic association studies of the circulating proteome as well as for EOC susceptibility may identify additional candidate biomarkers for EOC. Moreover, such studies may profile additional proteins (including CA-125, which was not profiled in the INTERVAL study) and include individuals of non-European ancestries, offering new opportunities for plasma protein biomarker discovery. The present study was unable to identify associations for EOC histotypes other than for the most common high-grade serous hisotype and this is another area where larger sample sizes might help. The INTERVAL and OCAC data sets used in this analysis were based on participants of European ancestry and there is a compelling need for similar trans-ancestry analyses. Smaller GWAS of ovarian cancer risk in women of African and East Asian ancestry have been reported by OCAC but there is no circulating protein level GWAS comparable to the INTERVAL study as yet in a cohort that is not of European ancestry [31,32]. A major strength of the current analysis was the ability to appraise the roles of over 1,300 proteins. A vital next step in assessing the role of the plasma proteome in EOC risk and validating our findings will involve directly measuring the 26 proteins shortlisted by our study in EOC case and control sample collections that have pre-diagnostic and longitudinal follow-up biospecimens available such as the Prostate, Lung, Colorectal and Ovarian

(PLCO) Cancer Screening Trial [33]. The pleiotropic associations observed at the 9q34.2 locus where eight of the ten SNPs associated with plasma protein levels were also associated with 62 other traits leaves open the possibility that some of these traits, rather than the protein levels, may underlie the association with EOC risk. Alternatively, some of these traits may lie up- or downstream of the protein levels and mediate the association with EOC risk as part of the same causal pathway. Further studies will be required to dissect these possibilities. For example, five of the eight SNPs at the 9q34.2 (*ABO*) locus are associated with low density lipoprotein-cholesterol (LDL-C) levels with the SNP alleles predicting lower LDL-C levels associating with reduced EOC risk (Table S3). This is consistent with a recent analysis based on the OCAC data set which showed that lower LDL-C level genetically predicted by SNPs in or near *HMGCR*, which encodes the enzyme inhibited by statins, was associated with reduced EOC risk [34].

In conclusion, our integrative analysis of large-scale proteomic and genomic data sets identified several associations between genetically predicted circulating protein levels and EOC risk that were statistically significant after FDR control and biologically plausible. These plasma proteins are candidate biomarkers with the potential for application in the early diagnosis of this aggressive gynecological cancer. The associations shed new light on EOC biology and should inform a range of follow-up laboratory-based studies and targeted biomarker validation projects wherein the 26 identified plasma proteins are directly tracked in incident EOC cases and controls over time.

Author Contribution statement

Conceptualization: PDPP, WZ, SPK. Data curation, Formal analysis, and Visualization: DPCC, GJ, XS, SPK. Project administration and Resources: JMS, WZ, PDPP, OCAC. Funding acquisition: JMS, PDPP, SPK. Supervision: PDPP, WZ, SPK. Writing - original draft: DPCC, GJ, XS. Writing - review & editing: All authors.

Conflict of Interest statement

The authors have no conflict of interest to declare.

Acknowledgments

This work was supported by a grant (R01 CA211574) from the US National Institutes of Health (National Cancer Institute). We thank the participants of the many studies contributing to the Ovarian Cancer Association Consortium (OCAC) and the INTERVAL study as well as the many researchers involved in OCAC and INTERVAL who have made this work possible. A full list of grant funding for OCAC is available in PMID: 28346442/PMCID: PMC5612337/DOI: 10.1038/ng.3826 and for INTERVAL is available in PMID: 29875488/PMCID: PMC6697541/DOI: 10.1038/s41586-018-0175-2.

Availability of data and code

The Ovarian Cancer Association Consortium (OCAC) data set can be downloaded from:
ftp://ftp.ebi.ac.uk/pub/databases/gwas/summary_statistics/PhelanCM_28346442_GCST004462

The genetic associations with circulating protein biomarkers were obtained from the INTERVAL study and can be downloaded from:
https://static-content.springer.com/esm/art%3A10.1038%2Fs41586-018-0175-2/MediaObjects/41586_2018_175_MOESM4_ESM.xlsx

Code used to perform the analysis reported in this paper is available at:
https://github.com/siddhartha-kar/circulating-proteins-and-ovarian-cancer

Other online tools used –

LDlink: https://ldlink.nci.nih.gov

Enrichr: https://amp.pharm.mssm.edu/Enrichr

WikiPathways: https://www.wikipathways.org

PhenoScanner: http://www.phenoscanner.medschl.cam.ac.uk/

References

[1] R.L. Siegel, K.D. Miller, A. Jemal, Cancer statistics, 2020, CA Cancer J Clin. 70 (2020) 7–30.
[2] F. Bray, J. Ferlay, I. Soerjomataram, R.L. Siegel, L.A. Torre, A. Jemal, Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries, CA Cancer J Clin. 68 (2018) 394–424.
[3] U.A. Matulonis, A.K. Sood, L. Fallowfield, B.E. Howitt, J. Sehouli, B.Y. Karlan, Ovarian cancer, Nat Rev Dis Primers. 2 (2016) 16061.
[4] R.C. Bast, M. Feeney, H. Lazarus, L.M. Nadler, R.B. Colvin, R.C. Knapp, Reactivity of a monoclonal antibody with human ovarian carcinoma, J. Clin. Invest. 68 (1981) 1331–1337.
[5] E.L. Moss, J. Hollingworth, T.M. Reynolds, The role of CA125 in clinical practice, J. Clin. Pathol. 58 (2005) 308–312.
[6] V. Dochez, H. Caillon, E. Vaucel, J. Dimet, N. Winer, G. Ducarme, Biomarkers and algorithms for diagnosis of ovarian cancer: CA125, HE4, RMI and ROMA, a review, J Ovarian Res. 12 (2019) 28.
[7] US Preventive Services Task Force, D.C. Grossman, S.J. Curry, D.K. Owens, M.J. Barry, K.W. Davidson, C.A. Doubeni, J.W. Epling, A.R. Kemper, A.H. Krist, A.E. Kurth, C.S. Landefeld, C.M. Mangione, M.G. Phipps, M. Silverstein, M.A. Simon, C.-W. Tseng, Screening for Ovarian Cancer: US Preventive Services Task Force Recommendation Statement, JAMA. 319 (2018) 588–594.

[8] R.G. Moore, D.S. McMeekin, A.K. Brown, P. DiSilvestro, M.C. Miller, W.J. Allard, W. Gajewski, R. Kurman, R.C. Bast, S.J. Skates, A novel multiple marker bioassay utilizing HE4 and CA125 for the prediction of ovarian cancer in patients with a pelvic mass, Gynecol Oncol. 112 (2009) 40–46.
[9] R.G. Moore, A.K. Brown, M.C. Miller, S. Skates, W.J. Allard, T. Verch, M. Steinhoff, G. Messerlian, P. DiSilvestro, C.O. Granai, R.C. Bast, The use of multiple novel tumor biomarkers for the detection of ovarian carcinoma in patients with a pelvic mass, Gynecol Oncol. 108 (2008) 402–408.
[10] S. Enroth, M. Berggrund, M. Lycke, J. Broberg, M. Lundberg, E. Assarsson, M. Olovsson, K. Stålberg, K. Sundfeldt, U. Gyllensten, High throughput proteomics identifies a high-accuracy 11 plasma protein biomarker signature for ovarian cancer, Commun Biol. 2 (2019) 221.
[11] S. Enroth, M. Berggrund, M. Lycke, M. Lundberg, E. Assarsson, M. Olovsson, K. Stålberg, K. Sundfeldt, U. Gyllensten, A two-step strategy for identification of plasma protein biomarkers for endometrial and ovarian cancer, Clin Proteomics. 15 (2018) 38.
[12] B.B. Sun, J.C. Maranville, J.E. Peters, D. Stacey, J.R. Staley, J. Blackshaw, S. Burgess, T. Jiang, E. Paige, P. Surendran, C. Oliver-Williams, M.A. Kamat, B.P. Prins, S.K. Wilcox, E.S. Zimmerman, A. Chi, N. Bansal, S.L. Spain, A.M. Wood, N.W. Morrell, J.R. Bradley, N. Janjic, D.J. Roberts, W.H. Ouwehand, J.A. Todd, N. Soranzo, K. Suhre, D.S. Paul, C.S. Fox, R.M. Plenge, J. Danesh, H. Runz, A.S. Butterworth, Genomic atlas of the human plasma proteome, Nature. 558 (2018) 73–79.
[13] C.M. Phelan, K.B. Kuchenbaecker, J.P. Tyrer, S.P. Kar, K. Lawrenson, S.J. Winham, J. Dennis, A. Pirie, M.J. Riggan, G. Chornokur, M.A. Earp, P.C. Lyra, J.M. Lee, S. Coetzee, J. Beesley, L. McGuffog, P. Soucy, E. Dicks, A. Lee, D. Barrowdale, J. Lecarpentier, G. Leslie, C.M. Aalfs, K.K.H. Aben, M. Adams, J. Adlard, I.L. Andrulis, H. Anton-Culver, N. Antonenkova, AOCS study group, G. Aravantinos, N. Arnold, B.K. Arun, B. Arver, J. Azzollini, J. Balmaña, S.N. Banerjee, L. Barjhoux, R.B. Barkardottir, Y. Bean, M.W. Beckmann, A. Beeghly-Fadiel, J. Benitez, M. Bermisheva, M.Q. Bernardini, M.J. Birrer, L. Bjorge, A. Black, K. Blankstein, M.J. Blok, C. Bodelon, N. Bogdanova, A. Bojesen, B. Bonanni, Å. Borg, A.R. Bradbury, J.D. Brenton, C. Brewer, L. Brinton, P. Broberg, A. Brooks-Wilson, F. Bruinsma, J. Brunet, B. Buecher, R. Butzow, S.S. Buys, T. Caldes, M.A. Caligo, I. Campbell, R. Cannioto, M.E. Carney, T. Cescon, S.B. Chan, J. Chang-Claude, S. Chanock, X.Q. Chen, Y.-E. Chiew, J. Chiquette, W.K. Chung, K.B.M. Claes, T. Conner, L.S. Cook, J. Cook, D.W. Cramer, J.M. Cunningham, A.A. D'Aloisio, M.B. Daly, F. Damiola, S.D. Damirovna, A. Dansonka-Mieszkowska, F. Dao, R. Davidson, A. DeFazio, C. Delnatte, K.F. Doheny, O. Diez, Y.C. Ding, J.A. Doherty, S.M. Domchek, C.M. Dorfling, T. Dörk, L. Dossus, M. Duran, M. Dürst, B. Dworniczak, D. Eccles, T. Edwards, R. Eeles, U. Eilber, B. Ejlertsen, A.B. Ekici, S. Ellis, M. Elvira, EMBRACE Study, K.H. Eng, C. Engel, D.G. Evans, P.A. Fasching, S. Ferguson, S.F. Ferrer, J.M. Flanagan, Z.C. Fogarty, R.T. Fortner, F. Fostira, W.D. Foulkes, G. Fountzilas, B.L. Fridley, T.M. Friebel, E. Friedman, D. Frost, P.A. Ganz, J. Garber, M.J. García, V. Garcia-Barberan, A. Gehrig, GEMO Study Collaborators, A. Gentry-Maharaj, A.-M. Gerdes, G.G. Giles, R. Glasspool, G. Glendon, A.K. Godwin, D.E. Goldgar, T. Goranova, M. Gore, M.H. Greene, J. Gronwald, S. Gruber, E. Hahnen, C.A. Haiman, N. Håkansson, U. Hamann, T.V.O. Hansen, P.A. Harrington, H.R. Harris, J. Hauke, HEBON Study, A. Hein, A. Henderson, M.A.T. Hildebrandt, P. Hillemanns, S. Hodgson, C.K. Høgdall, E. Høgdall, F.B.L. Hogervorst, H. Holland, M.J. Hooning, K. Hosking, R.-Y. Huang, P.J. Hulick, J. Hung, D.J. Hunter, D.G. Huntsman, T. Huzarski, E.N. Imyanitov, C. Isaacs, E.S. Iversen, L. Izatt, A. Izquierdo, A. Jakubowska, P. James, R. Janavicius, M. Jernetz, A. Jensen, U.B. Jensen, E.M. John, S. Johnatty, M.E. Jones, P. Kannisto, B.Y. Karlan, A. Karnezis, K. Kast, KConFab Investigators, C.J. Kennedy, E. Khusnutdinova, L.A. Kiemeney, J.I. Kiiski, S.-W. Kim, S.K. Kjaer, M. Köbel, R.K. Kopperud, T.A. Kruse, J. Kupryjanczyk, A. Kwong, Y. Laitman, D. Lambrechts, N. Larrañaga, M.C. Larson, C. Lazaro, N.D. Le, L. Le Marchand, J.W. Lee, S.B. Lele, A. Leminen, D. Leroux, J. Lester, F. Lesueur, D.A. Levine, D. Liang, C. Liebrich, J. Lilyquist, L. Lipworth, J. Lissowska, K.H. Lu, J. Lubiński, C. Luccarini, L. Lundvall, P.L. Mai, G. Mendoza-Fandiño, S. Manoukian, L.F.A.G. Massuger, T. May, S. Mazoyer, J.N. McAlpine, V. McGuire, J.R. McLaughlin, I. McNeish, H. Meijers-Heijboer, A.

Meindl, U. Menon, A.R. Mensenkamp, M.A. Merritt, R.L. Milne, G. Mitchell, F. Modugno, J. Moes-Sosnowska, M. Moffitt, M. Montagna, K.B. Moysich, A.M. Mulligan, J. Musinsky, K.L. Nathanson, L. Nedergaard, R.B. Ness, S.L. Neuhausen, H. Nevanlinna, D. Niederacher, R.L. Nussbaum, K. Odunsi, E. Olah, O.I. Olopade, H. Olsson, C. Olswold, D.M. O'Malley, K.-R. Ong, N.C. Onland-Moret, OPAL study group, N. Orr, S. Orsulic, A. Osorio, D. Palli, L. Papi, T.-W. Park-Simon, J. Paul, C.L. Pearce, I.S. Pedersen, P.H.M. Peeters, B. Peissel, A. Peixoto, T. Pejovic, L.M. Pelttari, J.B. Permuth, P. Peterlongo, L. Pezzani, G. Pfeiler, K.-A. Phillips, M. Piedmonte, M.C. Pike, A.M. Piskorz, S.R. Poblete, T. Pocza, E.M. Poole, B. Poppe, M.E. Porteous, F. Prieur, D. Prokofyeva, E. Pugh, M.A. Pujana, P. Pujol, P. Radice, J. Rantala, C. Rappaport-Fuerhauser, G. Rennert, K. Rhiem, P. Rice, A. Richardson, M. Robson, G.C. Rodriguez, C. Rodríguez-Antona, J. Romm, M.A. Rookus, M.A. Rossing, J.H. Rothstein, A. Rudolph, I.B. Runnebaum, H.B. Salvesen, D.P. Sandler, M.J. Schoemaker, L. Senter, V.W. Setiawan, G. Severi, P. Sharma, T. Shelford, N. Siddiqui, L.E. Side, W. Sieh, C.F. Singer, H. Sobol, H. Song, M.C. Southey, A.B. Spurdle, Z. Stadler, D. Steinemann, D. Stoppa-Lyonnet, L.E. Sucheston-Campbell, G. Sukiennicki, R. Sutphen, C. Sutter, A.J. Swerdlow, C.I. Szabo, L. Szafron, Y.Y. Tan, J.A. Taylor, M.-K. Tea, M.R. Teixeira, S.-H. Teo, K.L. Terry, P.J. Thompson, L.C.V. Thomsen, D.L. Thull, L. Tihomirova, A.V. Tinker, M. Tischkowitz, S. Tognazzo, A.E. Toland, A. Tone, B. Trabert, R.C. Travis, A. Trichopoulou, N. Tung, S.S. Tworoger, A.M. van Altena, D. Van Den Berg, A.H. van der Hout, R.B. van der Luijt, M. Van Heetvelde, E. Van Nieuwenhuysen, E.J. van Rensburg, A. Vanderstichele, R. Varon-Mateeva, A. Vega, D.V. Edwards, I. Vergote, R.A. Vierkant, J. Vijai, A. Vratimos, L. Walker, C. Walsh, D. Wand, S. Wang-Gohrke, B. Wappenschmidt, P.M. Webb, C.R. Weinberg, J.N. Weitzel, N. Wentzensen, A.S. Whittemore, J.T. Wijnen, L.R. Wilkens, A. Wolk, M. Woo, X. Wu, A.H. Wu, H. Yang, D. Yannoukakos, A. Ziogas, K.K. Zorn, S.A. Narod, D.F. Easton, C.I. Amos, J.M. Schildkraut, S.J. Ramus, L. Ottini, M.T. Goodman, S.K. Park, L.E. Kelemen, H.A. Risch, M. Thomassen, K. Offit, J. Simard, R.K. Schmutzler, D. Hazelett, A.N. Monteiro, F.J. Couch, A. Berchuck, G. Chenevix-Trench, E.L. Goode, T.A. Sellers, S.A. Gayther, A.C. Antoniou, P.D.P. Pharoah, Identification of 12 new susceptibility loci for different histotypes of epithelial ovarian cancer, Nat. Genet. 49 (2017) 680–691.

[14] E. Di Angelantonio, S.G. Thompson, S. Kaptoge, C. Moore, M. Walker, J. Armitage, W.H. Ouwehand, D.J. Roberts, J. Danesh, INTERVAL Trial Group, Efficiency and safety of varying the frequency of whole blood donation (INTERVAL): a randomised trial of 45 000 donors, Lancet. 390 (2017) 2360–2371.

[15] J.C. Rohloff, A.D. Gelinas, T.C. Jarvis, U.A. Ochsner, D.J. Schneider, L. Gold, N. Janjic, Nucleic Acid Ligands With Protein-like Side Chains: Modified Aptamers and Their Use as Diagnostic and Therapeutic Agents, Mol Ther Nucleic Acids. 3 (2014) e201.

[16] M.J. Machiela, S.J. Chanock, LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants, Bioinformatics. 31 (2015) 3555–3557.

[17] M.V. Kuleshov, M.R. Jones, A.D. Rouillard, N.F. Fernandez, Q. Duan, Z. Wang, S. Koplev, S.L. Jenkins, K.M. Jagodnik, A. Lachmann, M.G. McDermott, C.D. Monteiro, G.W. Gundersen, A. Ma'ayan, Enrichr: a comprehensive gene set enrichment analysis web server 2016 update, Nucleic Acids Res. 44 (2016) W90-97.

[18] D.N. Slenter, M. Kutmon, K. Hanspers, A. Riutta, J. Windsor, N. Nunes, J. Mélius, E. Cirillo, S.L. Coort, D. Digles, F. Ehrhart, P. Giesbertz, M. Kalafati, M. Martens, R. Miller, K. Nishida, L. Rieswijk, A. Waagmeester, L.M.T. Eijssen, C.T. Evelo, A.R. Pico, E.L. Willighagen, WikiPathways: a multifaceted pathway database bridging metabolomics to other omics research, Nucleic Acids Res. 46 (2018) D661–D667.

[19] M.A. Kamat, J.A. Blackshaw, R. Young, P. Surendran, S. Burgess, J. Danesh, A.S. Butterworth, J.R. Staley, PhenoScanner V2: an expanded tool for searching human genotype-phenotype associations, Bioinformatics. 35 (2019) 4851–4853.

[20] K.B. Kuchenbaecker, S.J. Ramus, J. Tyrer, A. Lee, H.C. Shen, J. Beesley, K. Lawrenson, L. McGuffog, S. Healey, J.M. Lee, T.J. Spindler, Y.G. Lin, T. Pejovic, Y. Bean, Q. Li, S. Coetzee, D. Hazelett, A. Miron, M. Southey, M.B. Terry, D.E. Goldgar, S.S. Buys, R. Janavicius, C.M. Dorfling, E.J. van Rensburg, S.L. Neuhausen, Y.C. Ding, T.V.O. Hansen, L. Jønson, A.-M. Gerdes, B. Ejlertsen, D. Barrowdale, J. Dennis, J. Benitez, A. Osorio, M.J. Garcia, I. Komenaka, J.N. Weitzel, P. Ganschow, P. Peterlongo, L. Bernard, A. Viel, B. Bonanni, B. Peissel, S. Manoukian, P. Radice, L. Papi, L. Ottini, F. Fostira, I. Konstantopoulou, J. Garber, D. Frost, J. Perkins, R. Platte, S. Ellis, EMBRACE, A.K. Godwin, R.K. Schmutzler, A. Meindl, C. Engel, C. Sutter, O.M. Sinilnikova, GEMO Study Collaborators, F. Damiola, S. Mazoyer, D. Stoppa-Lyonnet, K. Claes, K. De Leeneer, J. Kirk, G.C. Rodriguez, M. Piedmonte, D.M. O'Malley, M. de la Hoya, T. Caldes, K. Aittomäki, H. Nevanlinna, J.M. Collée, M.A. Rookus, J.C. Oosterwijk, Breast Cancer Family Registry, L. Tihomirova, N. Tung, U. Hamann, C. Isaccs, M. Tischkowitz, E.N. Imyanitov, M.A. Caligo, I.G. Campbell, F.B.L. Hogervorst, HEBON, E. Olah, O. Diez, I. Blanco, J. Brunet, C. Lazaro, M.A. Pujana, A. Jakubowska, J. Gronwald, J. Lubinski, G. Sukiennicki, R.B. Barkardottir, M. Plante, J. Simard, P. Soucy, M. Montagna, S. Tognazzo, M.R. Teixeira, KConFab Investigators, V.S. Pankratz, X. Wang, N. Lindor, C.I. Szabo, N. Kauff, J. Vijai, C.A. Aghajanian, G. Pfeiler, A. Berger, C.F. Singer, M.-K. Tea, C.M. Phelan, M.H. Greene, P.L. Mai, G. Rennert, A.M. Mulligan, S. Tchatchou, I.L. Andrulis, G. Glendon, A.E. Toland, U.B. Jensen, T.A. Kruse, M. Thomassen, A. Bojesen, J. Zidan, E. Friedman, Y. Laitman, M. Soller, A. Liljegren, B. Arver, Z. Einbeigi, M. Stenmark-Askmalm, O.I. Olopade, R.L. Nussbaum, T.R. Rebbeck, K.L. Nathanson, S.M. Domchek, K.H. Lu, B.Y. Karlan, C. Walsh, J. Lester, Australian Cancer Study (Ovarian Cancer Investigators), Australian Ovarian Cancer Study Group, A. Hein, A.B. Ekici, M.W. Beckmann, P.A. Fasching, D. Lambrechts, E. Van Nieuwenhuysen, I. Vergote, S. Lambrechts, E. Dicks, J.A. Doherty, K.G. Wicklund, M.A. Rossing, A. Rudolph, J. Chang-Claude, S. Wang-Gohrke, U. Eilber, K.B. Moysich, K. Odunsi, L. Sucheston, S. Lele, L.R. Wilkens, M.T. Goodman, P.J. Thompson, Y.B. Shvetsov, I.B. Runnebaum, M. Dürst, P. Hillemanns, T. Dörk, N. Antonenkova, N. Bogdanova, A. Leminen, L.M. Pelttari, R. Butzow, F. Modugno, J.L. Kelley, R.P. Edwards, R.B. Ness, A. du Bois, F. Heitz, I. Schwaab, P. Harter, K. Matsuo, S. Hosono, S. Orsulic, A. Jensen, S.K. Kjaer, E. Hogdall, H.N. Hasmad, M.A.N. Azmi, S.-H. Teo, Y.-L. Woo, B.L. Fridley, E.L. Goode, J.M. Cunningham, R.A. Vierkant, F. Bruinsma, G.G. Giles, D. Liang, M.A.T. Hildebrandt, X. Wu, D.A. Levine, M. Bisogna, A. Berchuck, E.S. Iversen, J.M. Schildkraut, P. Concannon, R.P. Weber, D.W. Cramer, K.L. Terry, E.M. Poole, S.S. Tworoger, E.V. Bandera, I. Orlow, S.H. Olson, C. Krakstad, H.B. Salvesen, I.L. Tangen, L. Bjorge, A.M. van Altena, K.K.H. Aben, L.A. Kiemeney, L.F.A.G. Massuger, M. Kellar, A. Brooks-Wilson, L.E. Kelemen, L.S. Cook, N.D. Le, C. Cybulski, H. Yang, J. Lissowska, L.A. Brinton, N. Wentzensen, C. Hogdall, L. Lundvall, L. Nedergaard, H. Baker, H. Song, D. Eccles, I. McNeish, J. Paul, K. Carty, N. Siddiqui, R. Glasspool, A.S. Whittemore, J.H. Rothstein, V. McGuire, W. Sieh, B.-T. Ji, W. Zheng, X.-O. Shu, Y.-T. Gao, B. Rosen, H.A. Risch, J.R. McLaughlin, S.A. Narod, A.N. Monteiro, A. Chen, H.-Y. Lin, J. Permuth-Wey, T.A. Sellers, Y.-Y. Tsai, Z. Chen, A. Ziogas, H. Anton-Culver, A. Gentry-Maharaj, U. Menon, P. Harrington, A.W. Lee, A.H. Wu, C.L. Pearce, G. Coetzee, M.C. Pike, A. Dansonka-Mieszkowska, A. Timorek, I.K. Rzepecka, J. Kupryjanczyk, M. Freedman, H. Noushmehr, D.F. Easton, K. Offit, F.J. Couch, S. Gayther, P.P. Pharoah, A.C. Antoniou, G. Chenevix-Trench, Consortium of Investigators of Modifiers of BRCA1 and BRCA2, Identification of six new susceptibility loci for invasive epithelial ovarian cancer, Nat. Genet. 47 (2015) 164–171.

[21] X. Wang, N.R. Tucker, G. Rizki, R. Mills, P.H. Krijger, E. de Wit, V. Subramanian, E. Bartell, X.-X. Nguyen, J. Ye, J. Leyton-Mange, E.V. Dolmatova, P. van der Harst, W. de Laat, P.T. Ellinor, C. Newton-Cheh, D.J. Milan, M. Kellis, L.A. Boyer, Discovery and validation of sub-threshold genome-wide association study loci using epigenomic signatures, Elife. 5 (2016).

[22] C. Porta, C. Paglino, A. Mosca, Targeting PI3K/Akt/mTOR Signaling in Cancer, Front Oncol. 4 (2014) 64.

[23] M.K. Ediriweera, K.H. Tennekoon, S.R. Samarakoon, Role of the PI3K/AKT/mTOR signaling pathway in ovarian cancer: Biological and therapeutic significance, Semin. Cancer Biol. 59 (2019) 147–160.
[24] C.-C. Huang, S.-H. Cheng, C.-H. Wu, W.-Y. Li, J.-S. Wang, M.-L. Kung, T.-H. Chu, S.-T. Huang, C.-T. Feng, S.-C. Huang, M.-H. Tai, Delta-like 1 homologue promotes tumorigenesis and epithelial-mesenchymal transition of ovarian high-grade serous carcinoma through activation of Notch signaling, Oncogene. 38 (2019) 3201–3215.
[25] L.-W. Yao, L.-L. Wu, L.-H. Zhang, W. Zhou, L. Wu, K. He, J.-C. Ren, Y.-C. Deng, D.-M. Yang, J. Wang, G.-G. Mu, M. Xu, J. Zhou, G.-A. Xiang, Q.-S. Ding, Y.-N. Yang, H.-G. Yu, MFAP2 is overexpressed in gastric cancer and promotes motility via the MFAP2/integrin α5β1/FAK/ERK pathway, Oncogenesis. 9 (2020) 17. https://doi.org/10.1038/s41389-020-0198-z.
[26] X. Gong, T. Dong, M. Niu, X. Liang, S. Sun, Y. Zhang, Y. Li, D. Li, lncRNA LCPAT1 Upregulation Promotes Breast Cancer Progression via Enhancing MFAP2 Transcription, Mol Ther Nucleic Acids. 21 (2020) 804–813.
[27] M. Wu, Y. Ding, X. Jiang, Y. Chen, N. Wu, L. Li, H. Wang, Y. Huang, N. Xu, L. Teng, Overexpressed MAGP1 Is Associated With a Poor Prognosis and Promotes Cell Migration and Invasion in Gastric Cancer, Front Oncol. 9 (2019) 1544.
[28] H. Arakawa, Netrin-1 and its receptors in tumorigenesis, Nat. Rev. Cancer. 4 (2004) 978–987.
[29] W. Hao, M. Yu, J. Lin, B. Liu, H. Xing, J. Yang, D. Sun, F. Chen, M. Jiang, C. Tang, X. Zhang, Y. Zhao, Y. Zhu, The pan-cancer landscape of netrin family reveals potential oncogenic biomarkers, Sci Rep. 10 (2020) 5224.
[30] A.D. Papanastasiou, G. Pampalakis, D. Katsaros, G. Sotiropoulou, Netrin-1 overexpression is predictive of ovarian malignancies, Oncotarget. 2 (2011) 363–367.
[31] A. Manichaikul, L.C. Peres, X.-Q. Wang, M.E. Barnard, D. Chyn, X. Sheng, Z. Du, J. Tyrer, J. Dennis, A.G. Schwartz, M.L. Cote, E. Peters, P.G. Moorman, M. Bondy, J.S. Barnholtz-Sloan, P. Terry, A.J. Alberg, E.V. Bandera, E. Funkhouser, A.H. Wu, C.L. Pearce, M. Pike, V.W. Setiawan, C.A. Haiman, African American Breast Cancer Consortium (AABC), African Ancestry Prostate Cancer Consortium (AAPC), J.R. Palmer, L. LeMarchand, L.R. Wilkens, A. Berchuck, J.A. Doherty, F. Modugno, R. Ness, K. Moysich, B.Y. Karlan, A.S. Whittemore, V. McGuire, W. Sieh, K. Lawrenson, S. Gayther, T.A. Sellers, P. Pharoah, J.M. Schildkraut, African American Cancer Epidemiology Study (AACES) and the Ovarian Cancer Association Consortium (OCAC), Identification of novel epithelial ovarian cancer loci in women of African ancestry, Int J Cancer. 146 (2020) 2987–2998.
[32] K. Lawrenson, F. Song, D.J. Hazelett, S.P. Kar, J. Tyrer, C.M. Phelan, R.I. Corona, N.I. Rodríguez-Malavé, J.-H. Seo, E. Adler, S.G. Coetzee, F. Segato, M.A.S. Fonseca, C.I. Amos, M.E. Carney, G. Chenevix-Trench, J. Choi, J.A. Doherty, W. Jia, G.J. Jin, B.-G. Kim, N.D. Le, J. Lee, L. Li, B.K. Lim, N.A. Adenan, M. Mizuno, B. Park, C.L. Pearce, K. Shan, Y. Shi, X.-O. Shu, W. Sieh, Australian Ovarian Cancer Study Group, P.J. Thompson, L.R. Wilkens, Q. Wei, Y.L. Woo, L. Yan, B.Y. Karlan, M.L. Freedman, H. Noushmehr, E.L. Goode, A. Berchuck, T.A. Sellers, S.-H. Teo, W. Zheng, K. Matsuo, S. Park, K. Chen, P.D.P. Pharoah, S.A. Gayther, M.T. Goodman, Genome-wide association studies identify susceptibility loci for epithelial ovarian cancer in east Asian women, Gynecol. Oncol. 153 (2019) 343–355.
[33] C.S. Zhu, W.-Y. Huang, P.F. Pinsky, C.D. Berg, M. Sherman, K.J. Yu, D.M. Carrick, A. Black, R. Hoover, P. Lenz, C. Williams, L. Hawkins, M. Chaloux, S. Yurgalevitch, S. Mathew, A. Miller, V. Olivo, A. Khan, S.M. Pretzel, D. Multerer, P. Beckmann, K.G. Broski, N.D. Freedman, The Prostate, Lung, Colorectal and Ovarian Cancer (PLCO) Screening Trial Pathology Tissue Resource, Cancer Epidemiol. Biomarkers Prev. 25 (2016) 1635–1642.
[34] J. Yarmolinsky, C.J. Bull, E.E. Vincent, J. Robinson, A. Walther, G.D. Smith, S.J. Lewis, C.L. Relton, R.M. Martin, Association Between Genetically Proxied Inhibition of HMG-CoA Reductase and Epithelial Ovarian Cancer, JAMA. 323 (2020) 646–655.

Table Legends

Table 1. Associations identified between genetically predicted circulating (plasma) protein levels and all invasive epithelial ovarian cancer risk.

Table 2. Pathways enriched among the genes encoding the 26 all invasive epithelial ovarian cancer risk-associated circulating protein biomarkers identified.

Figure Legends

Fig. 1. Regional genetic association plots. Genetic association with all invasive epithelial ovarian cancer risk (negative logarithm base 10 *P*-value) from the Ovarian Cancer Association Consortium study is plotted on the Y-axis and chromosomal position (build 37/hg 19) is plotted on the X-axis. SNPs are marked with blue dots or colored diamonds. SNPs marked with colored diamonds are the SNPs most strongly associated in the INTERVAL study with circulating (plasma) levels of the proteins named in the titles of the plots.

Tables

Table 1. Associations identified between genetically predicted circulating (plasma) protein levels and all invasive epithelial ovarian cancer risk.

Protein	Protein full name	Top SNP associated with plasma level of protein	Chr	Pos[a]	Cis /trans[b]	Gene encoding protein	Gene mapped to SNP	OR[c]	L95 %CL	U95 %CL	*P*	FDR
FA20B	Glycosaminoglycan xylosylkinase	rs587729126	9	136138765	trans	*FAM20B*	*ABO*	1.47	1.30	1.66	$8.1x10^{-10}$	$4.5x10^{-7}$
sICAM-2	Intercellular adhesion molecule 2	rs587729126	9	136138765	trans	*ICAM2*	*ABO*	0.65	0.57	0.75	$8.1x10^{-10}$	$4.5x10^{-7}$
VEGF sR2	Vascular endothelial growth factor receptor 2	rs635634	9	136155000	trans	*KDR*	*ABO*	0.84	0.80	0.89	$3.0x10^{-9}$	$4.5x10^{-7}$
ISLR2	Immunoglobulin superfamily containing leucine-rich repeat protein 2	rs115478735	9	136149711	trans	*ISLR2*	*ABO*	0.80	0.75	0.86	$2.8x10^{-9}$	$4.5x10^{-7}$
Met	Hepatocyte growth factor receptor	rs635634	9	136155000	trans	*MET*	*ABO*	0.77	0.71	0.84	$3.0x10^{-9}$	$4.5x10^{-7}$
TPST2	Protein-tyrosine sulfotransferase 2	rs115478735	9	136149711	trans	*TPST2*	*ABO*	1.36	1.23	1.51	$2.8x10^{-9}$	$4.5x10^{-7}$
LIF sR	Leukemia inhibitory factor receptor	rs635634	9	136155000	trans	*LIFR*	*ABO*	0.72	0.64	0.80	$3.0x10^{-9}$	$4.5x10^{-7}$
Endoglin	Endoglin	rs635634	9	136155000	trans	*ENG*	*ABO*	0.66	0.57	0.76	$3.0x10^{-9}$	$4.5x10^{-7}$
IGF-I sR	Insulin-like growth factor 1 receptor	rs635634	9	136155000	trans	*IGF1R*	*ABO*	0.64	0.55	0.74	$3.0x10^{-9}$	$4.5x10^{-7}$
sE-Selectin	E-selectin	rs2519093	9	136141870	trans	*SELE*	*ABO*	0.92	0.89	0.95	$4.2x10^{-9}$	$4.7x10^{-7}$
IL-3 Ra	Interleukin-3 receptor subunit alpha	rs2519093	9	136141870	trans	*IL3RA*	*ABO*	0.89	0.86	0.93	$4.2x10^{-9}$	$4.7x10^{-7}$
C1GLC	C1GALT1-specific chaperone 1	rs2519093	9	136141870	trans	*C1GALT1C1*	*ABO*	1.17	1.11	1.24	$4.2x10^{-9}$	$4.7x10^{-7}$
IR	Insulin receptor	rs507666	9	136149399	trans	*INSR*	*ABO*	0.85	0.80	0.90	$8.4x10^{-9}$	$8.6x10^{-7}$
QSOX2	Sulfhydryl oxidase 2	rs149092047	9	136139907	trans	*QSOX2*	*ABO*	1.09	1.06	1.13	$1.2x10^{-7}$	$1.0x10^{-5}$
FAM3D	Protein FAM3D	rs149092047	9	136139907	trans	*FAM3D*	*ABO*	1.12	1.07	1.16	$1.2x10^{-7}$	$1.0x10^{-5}$
GOLM1	Golgi membrane protein 1	rs149092047	9	136139907	trans	*GOLM1*	*ABO*	1.15	1.09	1.22	$1.2x10^{-7}$	$1.0x10^{-5}$
Desmoglein-2	Desmoglein-2	rs687621	9	136137065	trans	*DSG2*	*ABO*	1.39	1.23	1.57	$1.7x10^{-7}$	$1.3x10^{-5}$
ST4S6	Carbohydrate sulfotransferase 15	rs550057	9	136146597	trans	*CHST15*	*ABO*	0.79	0.72	0.86	$2.0x10^{-7}$	$1.4x10^{-5}$
Alkaline phosphatase, intestine	Intestinal-type alkaline phosphatase	rs550057	9	136146597	trans	*ALPI*	*ABO*	0.74	0.66	0.83	$2.0x10^{-7}$	$1.4x10^{-5}$
Coagulation Factor VIII	Coagulation Factor VIII	rs9411377	9	136145404	trans	*F8*	*ABO*	1.16	1.09	1.22	$5.5x10^{-7}$	$3.7x10^{-5}$
BGAT	Histo-blood group ABO system transferase	rs505922	9	136149229	cis	*ABO*	*ABO*	1.05	1.03	1.08	$6.7x10^{-7}$	$4.0x10^{-5}$
DC-SIGN	CD209 antigen	rs505922	9	136149229	trans	*CD209*	*ABO*	1.09	1.05	1.12	$6.7x10^{-7}$	$4.0x10^{-5}$
MFAP2	Microfibrillar-associated protein 2	rs4920605	1	17315425	cis	*MFAP2*	*MFAP2*	1.27	1.12	1.45	$1.8x10^{-4}$	0.011
SEMG2	Semenogelin-2	rs12881760	14	101176335	trans	*SEMG2*	*DLK1*	1.10	1.04	1.15	$4.0x10^{-4}$	0.021
DLK1	Protein delta homolog 1	rs12881760	14	101176335	cis	*DLK1*	*DLK1*	1.10	1.04	1.16	$4.0x10^{-4}$	0.021
NTNG1	Netrin-G1	rs115668827	1	107678268	cis	*NTNG1*	*NTNG1*	0.89	0.84	0.95	$4.9x10^{-4}$	0.025

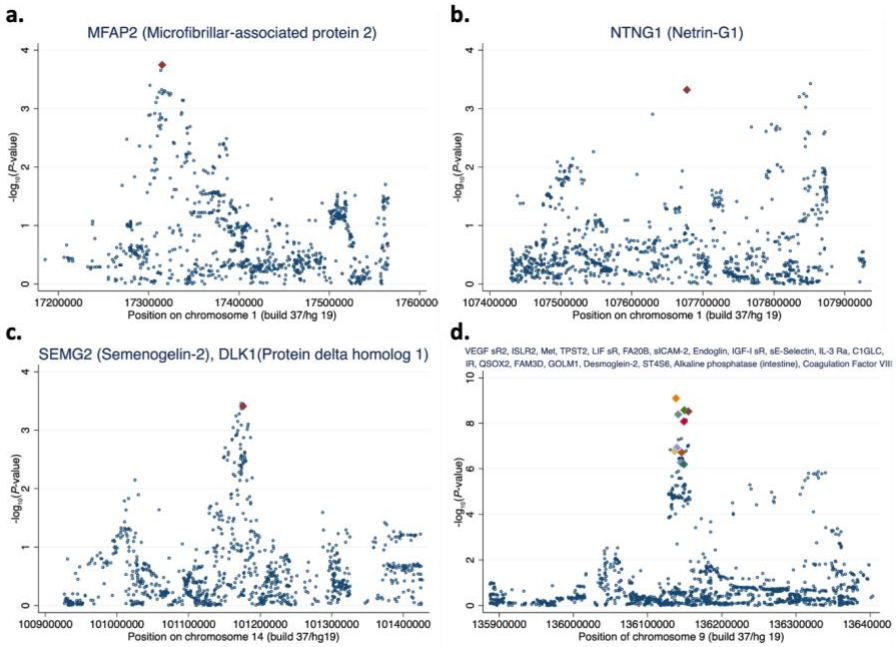
a Build 37/h19 position.
b *Cis* if the top SNP is < 1 Mb from the gene encoding protein and *trans* if the top SNP is > 1 Mb from the gene encoding the protein.
c Odds ratio (OR), lower 95% confidence limit (L95%CL), upper 95% confidence limit (U95%CL), P-value, and false discovery rate (FDR) from the current study. OR is in terms of all invasive EOC risk per standard deviation increase in circulating (plasma) protein level.

Table 2. Pathways enriched among the genes encoding the 26 all invasive epithelial ovarian cancer risk-associated circulating protein biomarkers identified.

Pathway[a]	Overlap[b]	*P*-value	Adjusted *P*[c]	Genes[d]
Pathways Regulating Hippo Signaling WP4540	4/98	7.5×10^{-6}	0.004	*INSR; KDR; MET; IGF1R*
Hippo-Merlin Signaling Dysregulation WP4541	4/120	1.7×10^{-5}	0.004	*INSR; KDR; MET; IGF1R*
Focal Adhesion-PI3K-Akt-mTOR-signaling pathway WP3932	5/303	3.9×10^{-5}	0.006	*INSR; IL3RA; KDR; MET; IGF1R*
PI3K-Akt Signaling Pathway WP4172	5/340	6.8×10^{-5}	0.008	*INSR; IL3RA; KDR; MET; IGF1R*
Ras Signaling WP4223	4/184	8.9×10^{-5}	0.008	*INSR; KDR; MET; IGF1R*
Ebola Virus Pathway on Host WP4217	3/129	6.1×10^{-4}	0.041	*CD209; ICAM2; IGF1R*
Canonical and Non-canonical Notch signaling WP3845	2/27	5.6×10^{-4}	0.044	*MFAP2; DLK1*

[a] From the "WikiPathways 2019 Human" pathway database (with associated WP identifier number).
[b] The number of genes out of the 26 genes evaluated/the total number of genes annotated to the pathway.
[c] Adjusted for testing 472 pathways.
[d] The genes (out of the 26 genes evaluated) that are annotated to the pathway.

Fig. 1


a.
MFAP2 (Microfibrillar-associated protein 2)
-log10(P-value)
Position on chromosome 1 (build 37/hg 19)
b.
NTNG1 (Netrin-G1)
-log10(P-value)
Position on chromosome 1 (build 37/hg 19)
c.
SEMG2 (Semenogelin-2), DLK1(Protein delta homolog 1)
-log10(P-value)
Position on chromosome 14 (build 37/hg19)
d.
VEGF sR2, ISLR2, Met, TPST2, LIF sR, FA20B, sICAM-2, Endoglin, IGF-I sR, sE-Selectin, IL-3 Ra, C1GLC,
IR, QSOX2, FAM3D, GOLM1, Desmoglein-2, ST4S6, Alkaline phosphatase (intestine), Coagulation Factor VIII
-log10(P-value)
Position of chromosome 9 (build 37/hg 19)

799