



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>

Peer reviewed version

Link to published version (if available):
[10.1016/j.ygyno.2020.11.016](https://doi.org/10.1016/j.ygyno.2020.11.016)

[Link to publication record in Explore Bristol Research](#)
PDF-document

This is the author accepted manuscript (AAM). The final published version (version of record) is available online via Elsevier at <https://doi.org/10.1016/j.ygyno.2020.11.016> . Please refer to any applicable terms of use of the publisher.

University of Bristol - Explore Bristol Research

General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:
<http://www.bristol.ac.uk/red/research-policy/pure/user-guides/ebr-terms/>

Genetically predicted circulating protein biomarkers and ovarian cancer risk

Daniel P. C. Considine*¹, Guochong Jia*², Xiang Shu^{2,3}, Joellen M. Schildkraut⁴, 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

53 **Abstract**

54

55 *Objective*

56

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

61

62 *Methods*

63

64 We used the germline genetic variants most strongly associated ($P < 1.5 \times 10^{-11}$) with plasma levels of
65 1,329 proteins in 3,301 healthy individuals from the INTERVAL study to predict circulating levels of
66 these proteins in 22,406 EOC cases and 40,941 controls from the Ovarian Cancer Association
67 Consortium (OCAC). Association testing was performed by weighting the beta coefficients and
68 standard errors for EOC risk from the OCAC study by the inverse of the beta coefficients from
69 INTERVAL.

70

71 *Results*

72

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

81

82 *Conclusion*

83

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

87

88 **Keywords**

89

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

92

93

94

95

96

97

98

99

100

101

102

103

104 **Research Highlights**

105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154

- 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

155 **Introduction**

156

157 Ovarian cancer is the most common cause of death from gynecological malignancy in the United
158 States and accounted for an estimated 295,000 incident cases and 184,000 deaths globally in 2018
159 [1,2]. Despite advances in treatment, survival rates in ovarian cancer continue to remain low, in
160 part, due to the late detection of most cases [3]. Nearly four decades after its discovery [4],
161 circulating levels of the protein cancer antigen 125 (CA-125) continue to be used to screen women at
162 high risk of developing ovarian cancer, such as those with a hereditary cancer syndrome, and
163 women with abnormal findings on examination and/or ultrasound. However, CA-125 has limited
164 sensitivity and specificity in these settings [5]. Furthermore, screening asymptomatic women for CA-
165 125 level, despite the use of serial measurements and algorithmic approaches to the interpretation
166 of these levels [6] – and even in combination with transvaginal ultrasound – does not reduce ovarian
167 cancer mortality and is not recommended by the US Preventive Services Task Force [7]. Human
168 epididymis secretory protein E4 (HE4) has been developed in recent years as a blood-based protein
169 biomarker for the diagnosis of ovarian carcinoma [8], and the combination of CA-125 and HE4 is a
170 more accurate predictor of ovarian malignancy than either biomarker alone [9]. However, there
171 remains an urgent unmet need to identify novel circulating protein biomarkers that will be more
172 useful for the early detection of this aggressive disease.

173

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

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

192

193 **Methods**

194

195 *Circulating (plasma) protein data set*

196

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

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

218

219 *Epithelial ovarian cancer data set*

220

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

232

233 *Statistical analysis*

234

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

255

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

259 same genomic region – a positional overlap that would reinforce the role of the SNP as a driver of
260 both circulating protein levels and ovarian cancer risk. We did this by visualizing ovarian cancer
261 genetic associations for all SNPs with MAF > 1% and imputation quality > 0.8 in the OCAC data set in
262 the 500 kb window centered on the top protein-associated SNP (i.e., +/- 250 kb on either side) using
263 two-way scatter plots generated in *Stata* (version 14, StataCorp LP, College Station, TX). For SNPs
264 with stronger *P*-values for association with ovarian cancer risk in OCAC as compared to the top
265 protein-associated SNP, the correlation between the stronger *P*-value SNPs and the top protein-
266 associated SNP was calculated using the LDlink online tool and data from the 1000 Genomes
267 European ancestry populations [16]. If the same SNP association signal drives both plasma protein
268 level and ovarian cancer risk, we expected one of the following three scenarios to be true: (i) the top
269 protein-associated SNP is also the top ovarian cancer associated SNP or (ii) it is strongly correlated
270 ($r^2 > 0.9$) with the top ovarian cancer associated SNP(s) or (iii) there are multiple independent ($r^2 <$
271 0.01) genetic association signals in OCAC in the same region and the top protein-associated SNP is
272 one of these associations. A second follow-up analysis of proteins that achieved FDR < 0.05 in our
273 study involved mapping these to the genes encoding them and evaluating the genes for enrichment
274 of pathways (at $P < 0.05$ after adjustment for testing multiple pathways) using the Enrichr online tool
275 [17] and the “WikiPathways Human 2019” database [18] that contains annotations for 472 known
276 biological pathways. A final follow-up analysis involved searching for genome-wide significant
277 associations ($P < 5 \times 10^{-8}$) between the top plasma protein level-associated SNP for each of the
278 proteins that achieved FDR < 0.05 in our study and other diseases and traits in the published (i.e.,
279 MEDLINE indexed) literature. This search was performed using the PhenoScanner (version 2) online
280 tool [19], querying published European-ancestry GWAS. The aim was to identify pleiotropic diseases
281 and traits that may provide an alternative explanation for the plasma protein-EOC risk associations
282 identified, stemming from their associations with the same top SNPs. Such pleiotropic diseases and
283 traits associated with the same SNPs may also be the cause or consequence of plasma protein level
284 changes that in turn are associated with EOC risk.

285

286 **Results**

287

288 Genetically predicted circulating levels of 26 proteins were associated with all invasive EOC risk at
289 FDR < 0.05 (13 positive and 13 inverse associations; Table 1 and Table S1). First, this included a
290 positive association between MFAP2 encoded by *MFAP2* on chromosome 1 and all invasive EOC risk
291 (Microfibrillar-associated protein 2; $P_{\text{Wald}} = 1.8 \times 10^{-4}$, FDR = 0.01). The top MFAP2 plasma protein-
292 associated SNP rs4920605 ($P_{\text{OCAC-GWAS-all}} = 1.82 \times 10^{-4}$) in the INTERVAL study was ~8 kb from the
293 transcription start site (TSS) of *MFAP2*. There was only one SNP in the same region, rs143483351
294 ($P_{\text{OCAC-GWAS-all}} = 1.76 \times 10^{-4}$), a multi-allelic variant 2 kb from rs4920605, with a slightly stronger
295 association with all invasive EOC risk (Fig. 1 (a) and Table S2). SNP rs143483351 could not be
296 evaluated in LDlink [16] for correlation with rs4920605 because it was a multi-allelic variant.
297 Second, our FDR < 0.05 results also included an inverse association between NTNG1 encoded by
298 *NTNG1* on chromosome 1 (in a genomic region distinct from *MFAP2*) and all invasive EOC risk
299 (Netrin-G1; $P_{\text{Wald}} = 4.9 \times 10^{-4}$, FDR = 0.03). The top NTNG1 plasma protein-associated SNP
300 rs115668827 ($P_{\text{OCAC-GWAS-all}} = 4.9 \times 10^{-4}$) in the INTERVAL study was ~4 kb from the TSS of *NTNG1*.
301 There was only one SNP in the same region, rs11185086 ($P_{\text{OCAC-GWAS-all}} = 3.8 \times 10^{-4}$), 173 kb from
302 rs115668827, with a stronger association with all invasive EOC risk (Fig. 1 (b) and Table S2).
303 However, rs11185086 and rs115668827 represented independent signals in the same region ($r^2 = 7 \times$
304 10^{-4}). Third, the list of 26 plasma proteins identified included positive associations between SEMG2
305 (Semenogelin-2) and ovarian cancer risk and DLK1 (Protein delta homolog 1) and ovarian cancer risk
306 (for both associations – $P_{\text{Wald}} = 4.0 \times 10^{-4}$, FDR = 0.02). The top SEMG2 plasma protein-associated
307 SNP and the top DLK1 plasma protein-associated SNP in the INTERVAL study was the same SNP,
308 rs12881760 ($P_{\text{OCAC-GWAS-all}} = 3.96 \times 10^{-4}$), which is ~16 kb from the TSS of *DLK1* on chromosome 14.
309 SEMG2 is encoded by *SEMG2* on chromosome 20 and rs12881760 is associated with its circulating
310 level by acting in *trans*. SNP rs12881760 is part of a cluster of three SNPs that includes rs10144381

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

314

315 The remaining 22 of the 26 all invasive EOC risk-associated circulating proteins identified were
316 proxied by 10 correlated SNPs ($r^2 > 0.38$) spanning a ~10 kb interval on chromosome 9 (Table 1). Three
317 of these SNPs were the top SNP for one protein each, four for two proteins each, two for three proteins
318 each, and one SNP was the top SNP for five proteins in the INTERVAL data set (Table 1). Ten proteins
319 demonstrated a positive association and 12 showed an inverse association with all invasive EOC risk.
320 Twenty-one of the 22 proteins were encoded by genes > 1 Mb away from this chromosome 9 interval
321 (*trans*-associations) and most were in fact encoded by genes located on other chromosomes. The
322 only exception to this was the plasma protein BGAT (Histo-blood group ABO system transferase)
323 encoded by *ABO* and the top BGAT plasma level-associated SNP, rs505922, is ~2 kb from the TSS of
324 *ABO*. The *ABO* locus (chromosome 9q34.2) is a known genome-wide significant ($P < 5 \times 10^{-8}$) locus for
325 all invasive and high-grade serous ovarian cancer risk [13,20]. The ten protein level-associated SNPs
326 spanned the *ABO* locus and were among the top 50 all invasive EOC risk SNPs in the 500 kb region
327 (Table S2 and Fig. 1 (d)). The top all invasive EOC risk SNP in the region, rs587729126 ($P_{\text{OCAC-GWAS-all}} =$
328 8.3×10^{-10}), was the top SNP in the INTERVAL study for association with circulating levels of FA20B
329 (Glycosaminoglycan xylosylkinase) and sICAM-2 (Intercellular adhesion molecule 2). This overlap of
330 top associations led to these two proteins emerging as the plasma proteins whose genetically
331 predicted levels were most strongly associated with all invasive EOC risk in our analysis (for both
332 associations – $P_{\text{Wald}} = 8.1 \times 10^{-10}$, FDR = 4.5×10^{-7} ; Table 1). The PhenoScanner search indicated that
333 eight of the ten protein level-associated SNPs that spanned the *ABO* locus were associated with 62
334 traits (Table S3). Overall, for all 26 proteins identified (associated with SNPs in the regions presented
335 in Fig. 1 and discussed above), we observed a clear overlap between the top circulating protein level-
336 associated SNP and the top all invasive EOC risk association, lending further confidence to the

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

344

345 **Discussion**

346

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

354

355 We observed that the top plasma protein level-associated SNP was either the top all invasive
356 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
357 of two independent ($r^2 < 0.01$) all invasive EOC risk associations in the region (Fig. 1 (b)) or it was part
358 of a cluster of highly correlated ($r^2 > 0.9$) SNPs that together marked the top all invasive EOC risk
359 association in the region (Fig. 1 (c)). This suggests that our results are unlikely to be due to linkage
360 disequilibrium contamination, i.e., regional genetic architecture where the top plasma protein level-
361 associated SNP is weakly correlated with the top all invasive EOC risk SNP and results in a spurious
362 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 ($P_{\text{pathway (adjusted)}} = 0.006$; 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 $P_{\text{pathway (adjusted)}} < 0.05$ was Notch signaling and this association was driven, in turn, by associations
388 between genetically predicted circulating levels of two Notch proteins, MFAP2 (chromosome

389 1p36.13) and DLK1 (chromosome 14q32.2), and all invasive EOC risk. It is noteworthy that these two
390 plasma proteins associated with EOC risk at FDR < 0.05 were encoded by genes located on distinct
391 chromosomes but the genes/proteins were members of the same biological pathway. DLK1, a non-
392 canonical Notch ligand, has a demonstrated role in promoting ovarian carcinogenesis via Notch
393 activation and epithelial-mesenchymal transition [24]. The microfibrillar-associated protein 2
394 (MFAP2), previously named microfibril-associated glycoprotein 1 (MAGP1), activates integrin
395 signaling and is a potential oncogene [25–27]. Another protein identified at FDR < 0.05, Netrin-G1 or
396 NTNG1, is involved in apoptosis and known to be dysregulated particularly in endocrine-related
397 tumors [28,29]. Further, the gene that encodes NTNG1 has been shown to be overexpressed in
398 malignant ovarian tumors [30].

399

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

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

426

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

434

435 **Author Contribution statement**

436

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

441

442 **Conflict of Interest statement**

443

444 The authors have no conflict of interest to declare.

445

446 **Acknowledgments**

447

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

454

455 **Availability of data and code**

456

457 The Ovarian Cancer Association Consortium (OCAC) data set can be downloaded from:

458 ftp://ftp.ebi.ac.uk/pub/databases/gwas/summary_statistics/PhelanCM_28346442_GCST004462

459

460 The genetic associations with circulating protein biomarkers were obtained from the INTERVAL study
461 and can be downloaded from:

462 [https://static-content.springer.com/esm/art%3A10.1038%2Fs41586-018-0175-
463 2/MediaObjects/41586_2018_175_MOESM4_ESM.xlsx](https://static-content.springer.com/esm/art%3A10.1038%2Fs41586-018-0175-2/MediaObjects/41586_2018_175_MOESM4_ESM.xlsx)

464

465 Code used to perform the analysis reported in this paper is available at:

466 <https://github.com/siddhartha-kar/circulating-proteins-and-ovarian-cancer>

467

468 Other online tools used –

469

470 LDlink: <https://ldlink.nci.nih.gov>

471

472 Enrichr: <https://amp.pharm.mssm.edu/Enrichr>

473

474 WikiPathways: <https://www.wikipathways.org>

475

476 PhenoScanner: <http://www.phenoscanter.medschl.cam.ac.uk/>

477

478 **References**

479

480 [1] R.L. Siegel, K.D. Miller, A. Jemal, Cancer statistics, 2020, *CA Cancer J Clin.* 70 (2020) 7–30.

481 [2] F. Bray, J. Ferlay, I. Soerjomataram, R.L. Siegel, L.A. Torre, A. Jemal, Global cancer statistics 2018:

482 GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries, *CA*
483 *Cancer J Clin.* 68 (2018) 394–424.

484 [3] U.A. Matulonis, A.K. Sood, L. Fallowfield, B.E. Howitt, J. Sehouli, B.Y. Karlan, Ovarian cancer, *Nat*
485 *Rev Dis Primers.* 2 (2016) 16061.

486 [4] R.C. Bast, M. Feeney, H. Lazarus, L.M. Nadler, R.B. Colvin, R.C. Knapp, Reactivity of a monoclonal
487 antibody with human ovarian carcinoma, *J. Clin. Invest.* 68 (1981) 1331–1337.

488 [5] E.L. Moss, J. Hollingworth, T.M. Reynolds, The role of CA125 in clinical practice, *J. Clin. Pathol.* 58
489 (2005) 308–312.

490 [6] V. Dochez, H. Caillon, E. Vaucel, J. Dimet, N. Winer, G. Ducarme, Biomarkers and algorithms for
491 diagnosis of ovarian cancer: CA125, HE4, RMI and ROMA, a review, *J Ovarian Res.* 12 (2019) 28.

492 [7] US Preventive Services Task Force, D.C. Grossman, S.J. Curry, D.K. Owens, M.J. Barry, K.W.

493 Davidson, C.A. Doubeni, J.W. Epling, A.R. Kemper, A.H. Krist, A.E. Kurth, C.S. Landefeld, C.M.

494 Mangione, M.G. Phipps, M. Silverstein, M.A. Simon, C.-W. Tseng, Screening for Ovarian Cancer:

495 US Preventive Services Task Force Recommendation Statement, *JAMA.* 319 (2018) 588–594.

- 496 [8] R.G. Moore, D.S. McMeekin, A.K. Brown, P. DiSilvestro, M.C. Miller, W.J. Allard, W. Gajewski, R.
497 Kurman, R.C. Bast, S.J. Skates, A novel multiple marker bioassay utilizing HE4 and CA125 for the
498 prediction of ovarian cancer in patients with a pelvic mass, *Gynecol Oncol.* 112 (2009) 40–46.
- 499 [9] R.G. Moore, A.K. Brown, M.C. Miller, S. Skates, W.J. Allard, T. Verch, M. Steinhoff, G. Messerlian,
500 P. DiSilvestro, C.O. Granai, R.C. Bast, The use of multiple novel tumor biomarkers for the
501 detection of ovarian carcinoma in patients with a pelvic mass, *Gynecol Oncol.* 108 (2008) 402–
502 408.
- 503 [10] S. Enroth, M. Berggrund, M. Lycke, J. Broberg, M. Lundberg, E. Assarsson, M. Olovsson, K.
504 Stålberg, K. Sundfeldt, U. Gyllensten, High throughput proteomics identifies a high-accuracy 11
505 plasma protein biomarker signature for ovarian cancer, *Commun Biol.* 2 (2019) 221.
- 506 [11] S. Enroth, M. Berggrund, M. Lycke, M. Lundberg, E. Assarsson, M. Olovsson, K. Stålberg, K.
507 Sundfeldt, U. Gyllensten, A two-step strategy for identification of plasma protein biomarkers
508 for endometrial and ovarian cancer, *Clin Proteomics.* 15 (2018) 38.
- 509 [12] B.B. Sun, J.C. Maranville, J.E. Peters, D. Stacey, J.R. Staley, J. Blackshaw, S. Burgess, T. Jiang, E.
510 Paige, P. Surendran, C. Oliver-Williams, M.A. Kamat, B.P. Prins, S.K. Wilcox, E.S. Zimmerman, A.
511 Chi, N. Bansal, S.L. Spain, A.M. Wood, N.W. Morrell, J.R. Bradley, N. Janjic, D.J. Roberts, W.H.
512 Ouwehand, J.A. Todd, N. Soranzo, K. Suhre, D.S. Paul, C.S. Fox, R.M. Plenge, J. Danesh, H. Runz,
513 A.S. Butterworth, Genomic atlas of the human plasma proteome, *Nature.* 558 (2018) 73–79.
- 514 [13] C.M. Phelan, K.B. Kuchenbaecker, J.P. Tyrer, S.P. Kar, K. Lawrenson, S.J. Winham, J. Dennis, A.
515 Pirie, M.J. Riggan, G. Chornokur, M.A. Earp, P.C. Lyra, J.M. Lee, S. Coetzee, J. Beesley, L.
516 McGuffog, P. Soucy, E. Dicks, A. Lee, D. Barrowdale, J. Lecarpentier, G. Leslie, C.M. Aalfs, K.K.H.
517 Aben, M. Adams, J. Adlard, I.L. Andrulis, H. Anton-Culver, N. Antonenkova, AOCs study group,
518 G. Aravantos, N. Arnold, B.K. Arun, B. Arver, J. Azzollini, J. Balmaña, S.N. Banerjee, L.
519 Barjhoux, R.B. Barkardottir, Y. Bean, M.W. Beckmann, A. Beeghly-Fadiel, J. Benitez, M.
520 Bermisheva, M.Q. Bernardini, M.J. Birrer, L. Bjorge, A. Black, K. Blankstein, M.J. Blok, C.
521 Bodelon, N. Bogdanova, A. Bojesen, B. Bonanni, Å. Borg, A.R. Bradbury, J.D. Brenton, C.
522 Brewer, L. Brinton, P. Broberg, A. Brooks-Wilson, F. Bruinsma, J. Brunet, B. Buecher, R. Butzow,
523 S.S. Buys, T. Caldes, M.A. Caligo, I. Campbell, R. Cannioto, M.E. Carney, T. Cescon, S.B. Chan, J.
524 Chang-Claude, S. Chanock, X.Q. Chen, Y.-E. Chiew, J. Chiquette, W.K. Chung, K.B.M. Claes, T.
525 Conner, L.S. Cook, J. Cook, D.W. Cramer, J.M. Cunningham, A.A. D’Aloisio, M.B. Daly, F.
526 Damiola, S.D. Damirovna, A. Dansonka-Mieszkowska, F. Dao, R. Davidson, A. DeFazio, C.
527 Delnatte, K.F. Doheny, O. Diez, Y.C. Ding, J.A. Doherty, S.M. Domchek, C.M. Dorfling, T. Dörk, L.
528 Dossus, M. Duran, M. Dürst, B. Dworniczak, D. Eccles, T. Edwards, R. Eeles, U. Eilber, B.
529 Ejlersen, A.B. Ekici, S. Ellis, M. Elvira, EMBRACE Study, K.H. Eng, C. Engel, D.G. Evans, P.A.
530 Fasching, S. Ferguson, S.F. Ferrer, J.M. Flanagan, Z.C. Fogarty, R.T. Fortner, F. Fostira, W.D.
531 Foulkes, G. Fountzilas, B.L. Fridley, T.M. Friebel, E. Friedman, D. Frost, P.A. Ganz, J. Garber, M.J.
532 García, V. Garcia-Barberan, A. Gehrig, GEMO Study Collaborators, A. Gentry-Maharaj, A.-M.
533 Gerdes, G.G. Giles, R. Glasspool, G. Glendon, A.K. Godwin, D.E. Goldgar, T. Goranova, M. Gore,
534 M.H. Greene, J. Gronwald, S. Gruber, E. Hahnen, C.A. Haiman, N. Håkansson, U. Hamann,
535 T.V.O. Hansen, P.A. Harrington, H.R. Harris, J. Hauke, HEBON Study, A. Hein, A. Henderson,
536 M.A.T. Hildebrandt, P. Hillemanns, S. Hodgson, C.K. Høgdall, E. Høgdall, F.B.L. Hogervorst, H.
537 Holland, M.J. Hooning, K. Hosking, R.-Y. Huang, P.J. Hulick, J. Hung, D.J. Hunter, D.G. Huntsman,
538 T. Huzarski, E.N. Imyanitov, C. Isaacs, E.S. Iversen, L. Izatt, A. Izquierdo, A. Jakubowska, P.
539 James, R. Janavicius, M. Jernetz, A. Jensen, U.B. Jensen, E.M. John, S. Johnatty, M.E. Jones, P.
540 Kannisto, B.Y. Karlan, A. Karnezis, K. Kast, KConFab Investigators, C.J. Kennedy, E.
541 Khusnutdinova, L.A. Kiemeny, J.I. Kiiski, S.-W. Kim, S.K. Kjaer, M. Köbel, R.K. Kopperud, T.A.
542 Kruse, J. Kupryjanczyk, A. Kwong, Y. Laitman, D. Lambrechts, N. Larrañaga, M.C. Larson, C.
543 Lazaro, N.D. Le, L. Le Marchand, J.W. Lee, S.B. Lele, A. Leminen, D. Leroux, J. Lester, F. Lesueur,
544 D.A. Levine, D. Liang, C. Liebrich, J. Lilyquist, L. Lipworth, J. Lissowska, K.H. Lu, J. Lubinowski, C.
545 Luccarini, L. Lundvall, P.L. Mai, G. Mendoza-Fandiño, S. Manoukian, L.F.A.G. Massuger, T. May,
546 S. Mazoyer, J.N. McAlpine, V. McGuire, J.R. McLaughlin, I. McNeish, H. Meijers-Heijboer, A.

547 Meindl, U. Menon, A.R. Mensenkamp, M.A. Merritt, R.L. Milne, G. Mitchell, F. Modugno, J.
 548 Moes-Sosnowska, M. Moffitt, M. Montagna, K.B. Moysich, A.M. Mulligan, J. Musinsky, K.L.
 549 Nathanson, L. Nedergaard, R.B. Ness, S.L. Neuhausen, H. Nevanlinna, D. Niederacher, R.L.
 550 Nussbaum, K. Odunsi, E. Olah, O.I. Olopade, H. Olsson, C. Olsword, D.M. O'Malley, K.-R. Ong,
 551 N.C. Onland-Moret, OPAL study group, N. Orr, S. Orsulic, A. Osorio, D. Palli, L. Papi, T.-W. Park-
 552 Simon, J. Paul, C.L. Pearce, I.S. Pedersen, P.H.M. Peeters, B. Peissel, A. Peixoto, T. Pejovic, L.M.
 553 Pelttari, J.B. Permut, P. Peterlongo, L. Pezzani, G. Pfeiler, K.-A. Phillips, M. Piedmonte, M.C.
 554 Pike, A.M. Piskorz, S.R. Poblete, T. Pocza, E.M. Poole, B. Poppe, M.E. Porteous, F. Prieur, D.
 555 Prokofyeva, E. Pugh, M.A. Pujana, P. Pujol, P. Radice, J. Rantala, C. Rappaport-Fuerhauser, G.
 556 Rennert, K. Rhiem, P. Rice, A. Richardson, M. Robson, G.C. Rodriguez, C. Rodríguez-Antona, J.
 557 Romm, M.A. Rookus, M.A. Rossing, J.H. Rothstein, A. Rudolph, I.B. Runnebaum, H.B. Salvesen,
 558 D.P. Sandler, M.J. Schoemaker, L. Senter, V.W. Setiawan, G. Severi, P. Sharma, T. Shelford, N.
 559 Siddiqui, L.E. Side, W. Sieh, C.F. Singer, H. Sobol, H. Song, M.C. Southey, A.B. Spurdle, Z. Stadler,
 560 D. Steinemann, D. Stoppa-Lyonnet, L.E. Sucheston-Campbell, G. Sukiennicki, R. Sutphen, C.
 561 Sutter, A.J. Swerdlow, C.I. Szabo, L. Szafron, Y.Y. Tan, J.A. Taylor, M.-K. Tea, M.R. Teixeira, S.-H.
 562 Teo, K.L. Terry, P.J. Thompson, L.C.V. Thomsen, D.L. Thull, L. Tihomirova, A.V. Tinker, M.
 563 Tischkowitz, S. Tognazzo, A.E. Toland, A. Tone, B. Trabert, R.C. Travis, A. Trichopoulou, N. Tung,
 564 S.S. Tworoger, A.M. van Altena, D. Van Den Berg, A.H. van der Hout, R.B. van der Lijst, M. Van
 565 Heetvelde, E. Van Nieuwenhuysen, E.J. van Rensburg, A. Vanderstichele, R. Varon-Mateeva, A.
 566 Vega, D.V. Edwards, I. Vergote, R.A. Vierkant, J. Vijai, A. Vratimos, L. Walker, C. Walsh, D.
 567 Wand, S. Wang-Gohrke, B. Wappenschmidt, P.M. Webb, C.R. Weinberg, J.N. Weitzel, N.
 568 Wentzensen, A.S. Whittemore, J.T. Wijnen, L.R. Wilkens, A. Wolk, M. Woo, X. Wu, A.H. Wu, H.
 569 Yang, D. Yannoukakos, A. Ziogas, K.K. Zorn, S.A. Narod, D.F. Easton, C.I. Amos, J.M. Schildkraut,
 570 S.J. Ramus, L. Ottini, M.T. Goodman, S.K. Park, L.E. Kelemen, H.A. Risch, M. Thomassen, K. Offit,
 571 J. Simard, R.K. Schmutzler, D. Hazelett, A.N. Monteiro, F.J. Couch, A. Berchuck, G. Chenevix-
 572 Trench, E.L. Goode, T.A. Sellers, S.A. Gayther, A.C. Antoniou, P.D.P. Pharoah, Identification of
 573 12 new susceptibility loci for different histotypes of epithelial ovarian cancer, *Nat. Genet.* 49
 574 (2017) 680–691.
 575 [14] E. Di Angelantonio, S.G. Thompson, S. Kaptoge, C. Moore, M. Walker, J. Armitage, W.H.
 576 Ouwehand, D.J. Roberts, J. Danesh, INTERVAL Trial Group, Efficiency and safety of varying the
 577 frequency of whole blood donation (INTERVAL): a randomised trial of 45 000 donors, *Lancet.*
 578 390 (2017) 2360–2371.
 579 [15] J.C. Rohloff, A.D. Gelinas, T.C. Jarvis, U.A. Ochsner, D.J. Schneider, L. Gold, N. Janjic, *Nucleic Acid*
 580 *Ligands With Protein-like Side Chains: Modified Aptamers and Their Use as Diagnostic and*
 581 *Therapeutic Agents*, *Mol Ther Nucleic Acids.* 3 (2014) e201.
 582 [16] M.J. Machiela, S.J. Chanock, LDlink: a web-based application for exploring population-specific
 583 haplotype structure and linking correlated alleles of possible functional variants,
 584 *Bioinformatics.* 31 (2015) 3555–3557.
 585 [17] M.V. Kuleshov, M.R. Jones, A.D. Rouillard, N.F. Fernandez, Q. Duan, Z. Wang, S. Koplev, S.L.
 586 Jenkins, K.M. Jagodnik, A. Lachmann, M.G. McDermott, C.D. Monteiro, G.W. Gundersen, A.
 587 Ma'ayan, Enrichr: a comprehensive gene set enrichment analysis web server 2016 update,
 588 *Nucleic Acids Res.* 44 (2016) W90-97.
 589 [18] D.N. Slenter, M. Kutmon, K. Hanspers, A. Riutta, J. Windsor, N. Nunes, J. Mélius, E. Cirillo, S.L.
 590 Coort, D. Digles, F. Ehrhart, P. Giesbertz, M. Kalafati, M. Martens, R. Miller, K. Nishida, L.
 591 Rieswijk, A. Waagmeester, L.M.T. Eijssen, C.T. Evelo, A.R. Pico, E.L. Willighagen, WikiPathways:
 592 a multifaceted pathway database bridging metabolomics to other omics research, *Nucleic*
 593 *Acids Res.* 46 (2018) D661–D667.
 594 [19] M.A. Kamat, J.A. Blackshaw, R. Young, P. Surendran, S. Burgess, J. Danesh, A.S. Butterworth, J.R.
 595 Staley, PhenoScanner V2: an expanded tool for searching human genotype-phenotype
 596 associations, *Bioinformatics.* 35 (2019) 4851–4853.

- 597 [20] K.B. Kuchenbaecker, S.J. Ramus, J. Tyrer, A. Lee, H.C. Shen, J. Beesley, K. Lawrenson, L.
598 McGuffog, S. Healey, J.M. Lee, T.J. Spindler, Y.G. Lin, T. Pejovic, Y. Bean, Q. Li, S. Coetzee, D.
599 Hazelett, A. Miron, M. Southey, M.B. Terry, D.E. Goldgar, S.S. Buys, R. Janavicius, C.M. Dorfling,
600 E.J. van Rensburg, S.L. Neuhausen, Y.C. Ding, T.V.O. Hansen, L. Jønson, A.-M. Gerdes, B.
601 Ejlertsen, D. Barrowdale, J. Dennis, J. Benitez, A. Osorio, M.J. Garcia, I. Komenaka, J.N. Weitzel,
602 P. Ganschow, P. Peterlongo, L. Bernard, A. Viel, B. Bonanni, B. Peissel, S. Manoukian, P. Radice,
603 L. Papi, L. Ottini, F. Fostira, I. Konstantopoulou, J. Garber, D. Frost, J. Perkins, R. Platte, S. Ellis,
604 EMBRACE, A.K. Godwin, R.K. Schmutzler, A. Meindl, C. Engel, C. Sutter, O.M. Sinilnikova, GEMO
605 Study Collaborators, F. Damiola, S. Mazoyer, D. Stoppa-Lyonnet, K. Claes, K. De Leeneer, J. Kirk,
606 G.C. Rodriguez, M. Piedmonte, D.M. O'Malley, M. de la Hoya, T. Caldes, K. Aittomäki, H.
607 Nevanlinna, J.M. Collée, M.A. Rookus, J.C. Oosterwijk, Breast Cancer Family Registry, L.
608 Tihomirova, N. Tung, U. Hamann, C. Isaccs, M. Tischkowitz, E.N. Imyanitov, M.A. Caligo, I.G.
609 Campbell, F.B.L. Hogervorst, HEBON, E. Olah, O. Diez, I. Blanco, J. Brunet, C. Lazaro, M.A.
610 Pujana, A. Jakubowska, J. Gronwald, J. Lubinski, G. Sukiennicki, R.B. Barkardottir, M. Plante, J.
611 Simard, P. Soucy, M. Montagna, S. Tognazzo, M.R. Teixeira, KConFab Investigators, V.S.
612 Pankratz, X. Wang, N. Lindor, C.I. Szabo, N. Kauff, J. Vijai, C.A. Aghajanian, G. Pfeiler, A. Berger,
613 C.F. Singer, M.-K. Tea, C.M. Phelan, M.H. Greene, P.L. Mai, G. Rennert, A.M. Mulligan, S.
614 Tchatchou, I.L. Andrulis, G. Glendon, A.E. Toland, U.B. Jensen, T.A. Kruse, M. Thomassen, A.
615 Bojesen, J. Zidan, E. Friedman, Y. Laitman, M. Soller, A. Liljegren, B. Arver, Z. Einbeigi, M.
616 Stenmark-Askmal, O.I. Olopade, R.L. Nussbaum, T.R. Rebbeck, K.L. Nathanson, S.M.
617 Domchek, K.H. Lu, B.Y. Karlan, C. Walsh, J. Lester, Australian Cancer Study (Ovarian Cancer
618 Investigators), Australian Ovarian Cancer Study Group, A. Hein, A.B. Ekici, M.W. Beckmann,
619 P.A. Fasching, D. Lambrechts, E. Van Nieuwenhuysen, I. Vergote, S. Lambrechts, E. Dicks, J.A.
620 Doherty, K.G. Wicklund, M.A. Rossing, A. Rudolph, J. Chang-Claude, S. Wang-Gohrke, U. Eilber,
621 K.B. Moysich, K. Odunsi, L. Sucheston, S. Lele, L.R. Wilkens, M.T. Goodman, P.J. Thompson, Y.B.
622 Shvetsov, I.B. Runnebaum, M. Dürst, P. Hillemanns, T. Dörk, N. Antonenkova, N. Bogdanova, A.
623 Leminen, L.M. Pelttari, R. Butzow, F. Modugno, J.L. Kelley, R.P. Edwards, R.B. Ness, A. du Bois,
624 F. Heitz, I. Schwaab, P. Harter, K. Matsuo, S. Hosono, S. Orsulic, A. Jensen, S.K. Kjaer, E. Hogdall,
625 H.N. Hasmad, M.A.N. Azmi, S.-H. Teo, Y.-L. Woo, B.L. Fridley, E.L. Goode, J.M. Cunningham, R.A.
626 Vierkant, F. Bruinsma, G.G. Giles, D. Liang, M.A.T. Hildebrandt, X. Wu, D.A. Levine, M. Bisogna,
627 A. Berchuck, E.S. Iversen, J.M. Schildkraut, P. Concannon, R.P. Weber, D.W. Cramer, K.L. Terry,
628 E.M. Poole, S.S. Tworoger, E.V. Bandera, I. Orlov, S.H. Olson, C. Krakstad, H.B. Salvesen, I.L.
629 Tangen, L. Borge, A.M. van Altena, K.K.H. Aben, L.A. Kiemenev, L.F.A.G. Massuger, M. Kellar, A.
630 Brooks-Wilson, L.E. Kelemen, L.S. Cook, N.D. Le, C. Cybulski, H. Yang, J. Lissowska, L.A. Brinton,
631 N. Wentzensen, C. Hogdall, L. Lundvall, L. Nedergaard, H. Baker, H. Song, D. Eccles, I. McNeish,
632 J. Paul, K. Carty, N. Siddiqui, R. Glasspool, A.S. Whittemore, J.H. Rothstein, V. McGuire, W. Sieh,
633 B.-T. Ji, W. Zheng, X.-O. Shu, Y.-T. Gao, B. Rosen, H.A. Risch, J.R. McLaughlin, S.A. Narod, A.N.
634 Monteiro, A. Chen, H.-Y. Lin, J. Permuth-Wey, T.A. Sellers, Y.-Y. Tsai, Z. Chen, A. Ziogas, H.
635 Anton-Culver, A. Gentry-Maharaj, U. Menon, P. Harrington, A.W. Lee, A.H. Wu, C.L. Pearce, G.
636 Coetzee, M.C. Pike, A. Dansonka-Mieszkowska, A. Timorek, I.K. Rzepecka, J. Kupryjanczyk, M.
637 Freedman, H. Noushmehr, D.F. Easton, K. Offit, F.J. Couch, S. Gayther, P.P. Pharoah, A.C.
638 Antoniou, G. Chenevix-Trench, Consortium of Investigators of Modifiers of BRCA1 and BRCA2,
639 Identification of six new susceptibility loci for invasive epithelial ovarian cancer, *Nat. Genet.* 47
640 (2015) 164–171.
- 641 [21] X. Wang, N.R. Tucker, G. Rizki, R. Mills, P.H. Krijger, E. de Wit, V. Subramanian, E. Bartell, X.-X.
642 Nguyen, J. Ye, J. Leyton-Mange, E.V. Dolmatova, P. van der Harst, W. de Laat, P.T. Ellinor, C.
643 Newton-Cheh, D.J. Milan, M. Kellis, L.A. Boyer, Discovery and validation of sub-threshold
644 genome-wide association study loci using epigenomic signatures, *Elife.* 5 (2016).
- 645 [22] C. Porta, C. Paglino, A. Mosca, Targeting PI3K/Akt/mTOR Signaling in Cancer, *Front Oncol.* 4
646 (2014) 64.

- 647 [23] M.K. Ediriweera, K.H. Tennekoon, S.R. Samarakoon, Role of the PI3K/AKT/mTOR signaling
648 pathway in ovarian cancer: Biological and therapeutic significance, *Semin. Cancer Biol.* 59
649 (2019) 147–160.
- 650 [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.
651 Feng, S.-C. Huang, M.-H. Tai, Delta-like 1 homologue promotes tumorigenesis and epithelial-
652 mesenchymal transition of ovarian high-grade serous carcinoma through activation of Notch
653 signaling, *Oncogene.* 38 (2019) 3201–3215.
- 654 [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.
655 Wang, G.-G. Mu, M. Xu, J. Zhou, G.-A. Xiang, Q.-S. Ding, Y.-N. Yang, H.-G. Yu, MFAP2 is
656 overexpressed in gastric cancer and promotes motility via the MFAP2/integrin $\alpha 5\beta 1$ /FAK/ERK
657 pathway, *Oncogenesis.* 9 (2020) 17. <https://doi.org/10.1038/s41389-020-0198-z>.
- 658 [26] X. Gong, T. Dong, M. Niu, X. Liang, S. Sun, Y. Zhang, Y. Li, D. Li, lncRNA LCPAT1 Upregulation
659 Promotes Breast Cancer Progression via Enhancing MFAP2 Transcription, *Mol Ther Nucleic*
660 *Acids.* 21 (2020) 804–813.
- 661 [27] M. Wu, Y. Ding, X. Jiang, Y. Chen, N. Wu, L. Li, H. Wang, Y. Huang, N. Xu, L. Teng, Overexpressed
662 MAGP1 Is Associated With a Poor Prognosis and Promotes Cell Migration and Invasion in
663 Gastric Cancer, *Front Oncol.* 9 (2019) 1544.
- 664 [28] H. Arakawa, Netrin-1 and its receptors in tumorigenesis, *Nat. Rev. Cancer.* 4 (2004) 978–987.
- 665 [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,
666 Y. Zhu, The pan-cancer landscape of netrin family reveals potential oncogenic biomarkers, *Sci*
667 *Rep.* 10 (2020) 5224.
- 668 [30] A.D. Papanastasiou, G. Pampalakis, D. Katsaros, G. Sotiropoulou, Netrin-1 overexpression is
669 predictive of ovarian malignancies, *Oncotarget.* 2 (2011) 363–367.
- 670 [31] A. Manichaikul, L.C. Peres, X.-Q. Wang, M.E. Barnard, D. Chyn, X. Sheng, Z. Du, J. Tyrer, J. Dennis,
671 A.G. Schwartz, M.L. Cote, E. Peters, P.G. Moorman, M. Bondy, J.S. Barnholtz-Sloan, P. Terry,
672 A.J. Alberg, E.V. Bandera, E. Funkhouser, A.H. Wu, C.L. Pearce, M. Pike, V.W. Setiawan, C.A.
673 Haiman, African American Breast Cancer Consortium (AABC), African Ancestry Prostate Cancer
674 Consortium (AAPC), J.R. Palmer, L. LeMarchand, L.R. Wilkens, A. Berchuck, J.A. Doherty, F.
675 Modugno, R. Ness, K. Moysich, B.Y. Karlan, A.S. Whittemore, V. McGuire, W. Sieh, K.
676 Lawrenson, S. Gayther, T.A. Sellers, P. Pharoah, J.M. Schildkraut, African American Cancer
677 Epidemiology Study (AACES) and the Ovarian Cancer Association Consortium (OCAC),
678 Identification of novel epithelial ovarian cancer loci in women of African ancestry, *Int J Cancer.*
679 146 (2020) 2987–2998.
- 680 [32] K. Lawrenson, F. Song, D.J. Hazelett, S.P. Kar, J. Tyrer, C.M. Phelan, R.I. Corona, N.I. Rodríguez-
681 Malavé, J.-H. Seo, E. Adler, S.G. Coetzee, F. Segato, M.A.S. Fonseca, C.I. Amos, M.E. Carney, G.
682 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,
683 N.A. Adenan, M. Mizuno, B. Park, C.L. Pearce, K. Shan, Y. Shi, X.-O. Shu, W. Sieh, Australian
684 Ovarian Cancer Study Group, P.J. Thompson, L.R. Wilkens, Q. Wei, Y.L. Woo, L. Yan, B.Y. Karlan,
685 M.L. Freedman, H. Noushmehr, E.L. Goode, A. Berchuck, T.A. Sellers, S.-H. Teo, W. Zheng, K.
686 Matsuo, S. Park, K. Chen, P.D.P. Pharoah, S.A. Gayther, M.T. Goodman, Genome-wide
687 association studies identify susceptibility loci for epithelial ovarian cancer in east Asian women,
688 *Gynecol. Oncol.* 153 (2019) 343–355.
- 689 [33] C.S. Zhu, W.-Y. Huang, P.F. Pinsky, C.D. Berg, M. Sherman, K.J. Yu, D.M. Carrick, A. Black, R.
690 Hoover, P. Lenz, C. Williams, L. Hawkins, M. Chaloux, S. Yurgalevitch, S. Mathew, A. Miller, V.
691 Olivo, A. Khan, S.M. Pretzel, D. Multerer, P. Beckmann, K.G. Broski, N.D. Freedman, The
692 Prostate, Lung, Colorectal and Ovarian Cancer (PLCO) Screening Trial Pathology Tissue
693 Resource, *Cancer Epidemiol. Biomarkers Prev.* 25 (2016) 1635–1642.
- 694 [34] J. Yarmolinsky, C.J. Bull, E.E. Vincent, J. Robinson, A. Walther, G.D. Smith, S.J. Lewis, C.L. Relton,
695 R.M. Martin, Association Between Genetically Proxied Inhibition of HMG-CoA Reductase and
696 Epithelial Ovarian Cancer, *JAMA.* 323 (2020) 646–655.
- 697

698 **Table Legends**

699

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

702

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

705

706 **Figure Legends**

707

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

714

715

716

717

718

719

720

721

722

723

724

725

726

727

728

729

730

731

732

733

734

735

736

737

738

739

740

741

742

743

744

745

746

747

748

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-1 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 x 10 ⁻⁶	0.004	<i>INSR; KDR; MET; IGF1R</i>
Hippo-Merlin Signaling Dysregulation WP4541	4/120	1.7 x 10 ⁻⁵	0.004	<i>INSR; KDR; MET; IGF1R</i>
Focal Adhesion-PI3K-Akt-mTOR-signaling pathway WP3932	5/303	3.9 x 10 ⁻⁵	0.006	<i>INSR; IL3RA; KDR; MET; IGF1R</i>
PI3K-Akt Signaling Pathway WP4172	5/340	6.8 x 10 ⁻⁵	0.008	<i>INSR; IL3RA; KDR; MET; IGF1R</i>
Ras Signaling WP4223	4/184	8.9 x 10 ⁻⁵	0.008	<i>INSR; KDR; MET; IGF1R</i>
Ebola Virus Pathway on Host WP4217	3/129	6.1 x 10 ⁻⁴	0.041	<i>CD209; ICAM2; IGF1R</i>
Canonical and Non-canonical Notch signaling WP3845	2/27	5.6 x 10 ⁻⁴	0.044	<i>MFAP2; DLK1</i>

^a From the “WikiPathways 2019 Human” pathway database (with associated WP identifier number).

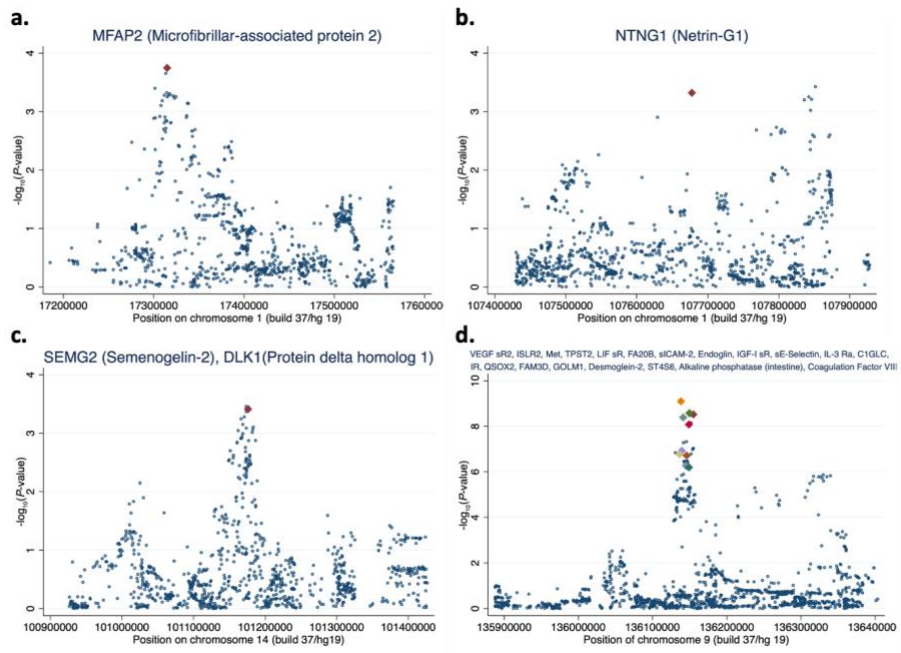
^b The number of genes out of the 26 genes evaluated/the total number of genes annotated to the pathway.

^c Adjusted for testing 472 pathways.

^d The genes (out of the 26 genes evaluated) that are annotated to the pathway.

762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798

Fig. 1



799