

RESEARCH ARTICLE

The blood transcriptome prior to ovarian cancer diagnosis: A case-control study in the NOWAC postgenome cohort

Mie Jareid^{1*}, Igor Snapkov², Marit Holden³, Lill-Tove Rasmussen Busund⁴, Eiliv Lund^{1,5}, Therese Haugdahl Nøst¹

1 Faculty of Health Sciences, Department of Community Medicine, UiT – The Arctic University of Norway, Tromsø, Norway, **2** Department of Immunology, Oslo University Hospital Rikshospitalet and University of Oslo, Oslo, Norway, **3** Norwegian Computing Center, Oslo, Norway, **4** Faculty of Health Sciences, Department of Medical Biology, UiT – The Arctic University of Norway, Tromsø, Norway, **5** Cancer Registry of Norway, Oslo, Norway

* mie.jareid@uit.no



OPEN ACCESS

Citation: Jareid M, Snapkov I, Holden M, Busund L-TR, Lund E, Nøst TH (2021) The blood transcriptome prior to ovarian cancer diagnosis: A case-control study in the NOWAC postgenome cohort. PLoS ONE 16(8): e0256442. <https://doi.org/10.1371/journal.pone.0256442>

Editor: Elizabeth Christie, Peter MacCallum Cancer Institute, AUSTRALIA

Received: February 24, 2021

Accepted: August 6, 2021

Published: August 27, 2021

Copyright: © 2021 Jareid et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: Data cannot be shared publicly because of national and institutional policies on research ethics and data security. Access to data requires approval from the Regional Committee for Medical and Health Research Ethics (REC North) and is conditional on adherence to the procedures of the Norwegian Women and Cancer Study and UiT – The Arctic University of Norway. Please contact the Norwegian Women and Cancer Study, att. Tonje Braaten <tonje.braaten@uit.no> or Arne Bastian Wiik <arne.b.wiik@uit.no>.

Abstract

Epithelial ovarian cancer (EOC) has a 5-year relative survival of 50%, partly because markers of early-stage disease are not available in current clinical diagnostics. The aim of the present study was to investigate whether EOC is associated with transcriptional profiles in blood collected up to 7 years before diagnosis. For this, we used RNA-stabilized whole blood, which contains circulating immune cells, from a sample of EOC cases from the population-based Norwegian Women and Cancer (NOWAC) postgenome cohort. We explored case-control differences in gene expression in all EOC (66 case-control pairs), as well as associations between gene expression and metastatic EOC (56 pairs), serous EOC (45 pairs, 44 of which were metastatic), and interval from blood sample collection to diagnosis (≤ 3 or > 3 years; 34 and 31 pairs, respectively). Lastly, we assessed differential expression of genes associated with EOC in published functional genomics studies that used blood samples collected from newly diagnosed women. After adjustment for multiple testing, this nested case-control study revealed no significant case-control differences in gene expression in all EOC (false discovery rate $q > 0.96$). With the exception of a few probes, the \log_2 fold change values obtained in gene-wise linear models were below ± 0.2 . P-values were lowest in analyses of metastatic EOC (80% of which were serous EOC). No common transcriptional profile was indicated by interval to diagnosis; when comparing the 100 genes with the lowest p-values in gene-wise tests in samples collected ≤ 3 and > 3 years before EOC diagnosis, no overlap in these genes was observed. Among 86 genes linked to ovarian cancer in previous publications, our data contained expression values for 42, and of these, tests of *LIME1*, *GPR162*, *STAB1*, and *SKAP1*, resulted in unadjusted $p < 0.05$. Although limited by sample size, our findings indicated less variation in blood gene expression between women with similar tumor characteristics.

Funding: The microarray analyses were financed by a grant awarded to EL (ERC-AdG 11 232997 TICE) by the European Research Council (<https://erc.europa.eu/>). Open access publication of this article was financed by a grant awarded to MJ by the UiT publication fund (<https://en.uit.no/publishing>) The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Introduction

Epithelial ovarian cancer (EOC) is the eighth most common cancer among Norwegian women, who have a 1.3% risk of developing this cancer by the age of 75 years. Further, age-standardized rates show that EOC is the fifth most common cause of cancer death [1]. EOC is often diagnosed in late stages, with 70% of cases diagnosed with stage III or IV disease. This is partly because markers of early-stage disease are not available in current clinical diagnostics. The symptoms that could lead to EOC diagnosis tend to manifest only after metastasis has already occurred, at which point curative treatment is difficult to achieve. The most common EOC subtype, serous carcinoma, is associated with a particularly poor prognosis [2].

The origin and pathogenesis of EOC vary by subtype, and are still not completely understood. Models have suggested that serous tumors exist as *in-situ* or stage I or II invasive tumors for a median of 5.1 years (95% confidence interval [CI]: 3.2–8.1 years), and advancement to stage III or IV can occur up to 2 years (median 0.8, 95% CI: 0.4–1.9 years) before diagnosis [3].

Ovarian malignancies are associated with skewed proportions of circulating immune cell types, and immunologic studies suggest induction of tumor tolerance through local, and potentially also systemic, immunosuppression mechanisms [4]. Functional genomic studies of circulating immune cells collected at EOC diagnosis have identified markers of risk, presence of tumor in patients, or prognosis [5–13]. However, few have investigated the blood transcriptome [8,10–13].

Whereas blood collected postdiagnostically reflects clinical cancer, random sampling of the general population allows researchers to study persons at different prediagnostic stages of tumorigenesis [14]. The aim of the present study was to investigate whether EOC is associated with transcriptional profiles in blood collected up to 7 years before diagnosis. For this, we used RNA-stabilized whole blood, which contains circulating immune cells, from a sample of EOC cases from the population-based Norwegian Women and Cancer (NOWAC) postgenome cohort. We explored case-control differences in gene expression in all EOC, as well as associations between gene expression and metastatic EOC, serous EOC, and interval from blood sample collection to diagnosis (≤ 3 or > 3 years). Lastly, we assessed differential expression of genes associated with EOC in published functional genomics studies that used blood samples collected from newly diagnosed women.

Materials and methods

Study population and sample collection

The present case-control study was nested within the NOWAC postgenome cohort, a subcohort of the NOWAC Study [15]. The NOWAC postgenome cohort is a population-based, prospective study initiated with the purpose of exploring associations between blood gene expression and cancer, with the inclusion of questionnaire information on a variety of exposures and lifestyle factors. Participants were recruited to the NOWAC Study by mail; those who consented to donate blood received a sampling kit with PAXgene blood collection tubes with RNA-preserving buffer (Preanalytix GmbH, Hembrechtikon, Switzerland). Participants then took this kit to a general practitioner's office, where the blood sample was collected. Between 2003 and 2006, blood samples from close to 50,000 women born between 1943 and 1957 were collected [16] and shipped to the study center, where they were stored at -80 degrees Celsius between 24 hours and 3 days after their collection.

Case ascertainment and assignment of matched controls. Epithelial ovarian cancer cases were identified through linkage to the Cancer Registry of Norway using the personal identification number assigned to all Norwegian citizens and permanent residents. Norwegian

health care providers are obligated to report all cancer cases to the registry, which ensures near complete national follow-up [17]. Participants of the NOWAC postgenome cohort with registered cancer of the ovary or fallopian tube (International Classification of Diseases revision 7, location 175) diagnosed between April 2004 and April 2011 ($n = 95$) were eligible for inclusion in this analysis. Tumors were then categorized as borderline, non-epithelial, EOC, and serous EOC; metastasis status was categorized as none, any, or unknown. Controls were matched to cases by birth year and blood sample storage time.

Questionnaire variables. On the day of blood sample collection, participants completed a two-page questionnaire concerning recent exposures. Information on variables known to be associated with EOC risk [18] and with gene expression in leukocytes was extracted from this questionnaire, and from NOWAC Study questionnaires: body mass index (BMI) [19], current smoking [20] including number of cigarettes smoked, parity [21], menopausal status [22], and current hormone replacement therapy (HRT) use [23]. We also included current oral contraceptive (OC) use, which modulates EOC risk and could influence gene expression.

Sample processing

Blood samples were processed at the Genomics Core Facility at the Norwegian University of Science and Technology according to the protocols of kit manufacturers. Samples from case-control pairs were processed together, blinded for case/control status. Total RNA was extracted from whole blood using the PAXgene Blood miRNA Kit (Qiagen GmbH, Hombrechtikon, Switzerland), and cRNA was prepared with the Illumina TotalPrep-96 RNA Amplification Kit (Ambion Inc., Austin, TX, USA). RNA quantity and purity were assessed using a NanoDrop ND 8000 spectrophotometer (ThermoFisher Scientific, Wilmington, DE, USA), and RNA integrity was assessed using Bioanalyzer capillary electrophoresis (Agilent Technologies, Palo Alto, CA, USA). cRNA was hybridized to Illumina HumanHT-12 v4 Expression BeadChip microarrays (Illumina, Inc. San Diego, CA, USA). Illumina GenomeStudio software was used to extract the raw data.

Preprocessing of microarray data. Background correction was performed using negative control probes (limma package, *nec* function) [24]. Probes reported by Illumina to be of poor quality, that were not annotated, that had a detection p -value < 0.05 , or that were present in less than 70% of the samples, were filtered out. Quantile normalization (*lumi*, *LumiN* function) [25] and \log_2 transformation (*lumi*, *LumiT*) was performed on the expression values. Finally, probes were mapped and annotated (*lumi*, *nuID2RefSeqID* and *illuminaHumanv4.db*). If multiple probes mapped to the same gene, all were kept in the dataset as duplicates/triplicates.

Statistical analysis

Preliminary quality control of laboratory measurements resulted in the exclusion of five case-control pairs; therefore 90 case-control pairs were included in the preprocessing of microarray data. After preprocessing, the dataset included 12,153 probes for 9,633 genes across 90 cases and 90 controls. We then further excluded cases with borderline tumors (20 pairs) and non-epithelial tumors (4 pairs), leaving 66 case-control pairs in the final dataset. We assessed case-control differences in gene expression in all EOC (66 pairs), as well as associations between gene expression and metastatic EOC (56 pairs), serous EOC (45 pairs, 44 of which were metastatic), and interval from blood sample collection to diagnosis (≤ 3 years and > 3 years, 34 pairs and 31 pairs, respectively). Exclusions and analytical samples are shown in Fig 1. To protect the identity of participants, date of diagnosis was generalized to the month of diagnosis. This resulted in negative follow-up time for one case, and exclusion of this case-control pair from the analysis of blood samples collected ≤ 3 years before diagnosis. The questionnaire variables

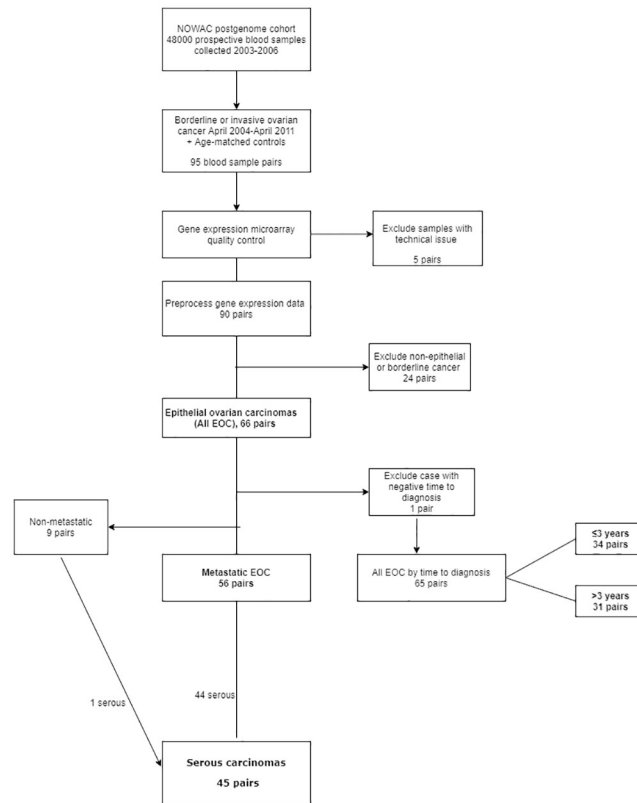


Fig 1. Flow chart of exclusions and analytic groups in gene expression analyses. Bold text indicates analyzed groups. The group “All EOC by time to diagnosis” was used in the analysis of all EOC adjusted for leukocyte populations.

<https://doi.org/10.1371/journal.pone.0256442.g001>

BMI ($<$ median 25.8, \geq 25.8), current smoking (yes/no), parity (0, 1–2, 3–4, \geq 5), menopausal status (pre- or perimenopausal, postmenopausal), current HRT use (yes/no), current OC use (yes/no) were evaluated as potential confounders by testing their association with case status by two-sided t-tests or chi-square tests. Further, to facilitate the evaluation of confounding by differences in leukocyte populations between cases and controls, we estimated leukocyte fractions in individual samples based on gene expression values, using the cell deconvolution procedure Cibersort and the LM22 signature matrix [26]. Variables that were associated with both case/control status ($p < 0.1$) and gene expression (global test [see below], family-wise error-rate adjusted $p < 0.05$) were adjusted for in the analyses.

Assessment of case-control differences in gene expression

To explore case-control differences in gene expression in all EOC, we computed a dissimilarity matrix with Euclidean distances and applied hierarchical clustering using Ward’s method to create a dendrogram. Further, we made a multidimensional scaling plot to display distances between samples. We used the global test [27] to assess case/control sample status as a function of difference in overall gene expression in all EOC, metastatic EOC, and serous EOC. Using linear models in the limma package [24], we assessed differences in expression of single genes (\log_2 fold change [FC] values) between cases and matched controls in all EOC, metastatic EOC, serous EOC, and EOC cases diagnosed \leq 3 years and $>$ 3 years after blood sample collection.

We used the global test [27] to evaluate associations between potential confounding variables and gene expression overall, and created an adjusted gene-wise model of all EOC. To explore expression differences in sets of genes, we used the mroast method (using 10^5 rotations) [28] to test gene sets from the C2, C5, C7 [29,30] and KEGG [31] collections in the Broad Institute databases [32].

Genes were considered differentially expressed if the false discovery rate (FDR)-adjusted p-value (q value) was <0.05 . We present non-FDR-adjusted p-values in the tables and text. The open source softwares R [3.1.2 and 3.2.1] (Vienna, Austria; www.r-project.org) and Bioconductor (bioconductor.org) were used for the analyses, with the exception of the chi-square test [33].

Gene Ontology enrichment. To explore the biological functions of the genes indicated according to case-control differences in expression, we used the R package clusterProfiler v.3.12.0 [34], which assesses potential overrepresentation of Gene Ontology (GO) terms [35,36]. We assessed the 100 probes with the lowest p-values in the limma models without covariate adjustments.

Differential expression of genes identified in published functional genomics studies. We used the metastatic EOC group to test case-control differences in the expression of seven sets of 5–33 genes reported to be associated with EOC in published functional genomics studies that used blood samples collected from newly diagnosed women. Of these, two gene sets were identified in whole blood gene expression studies comparing patients given a poor or better prognosis according to tumor characteristics [10,11]. Five gene sets were identified in case-control studies of DNA methylation in circulating leukocytes. We tested for differential expression of genes adjacent to CpG sites where differential methylation was reported indicative of EOC case status [6,9], CpGs indicative of EOC predisposition [5]; CpGs where methylation mediates genetic risk of EOC [7], and a set of genes where expression levels was suggested to mediate genotype-associated risk of EOC [8]. We tested a total 86 genes using a two-sided t-test for each gene, and did not adjust the p-values for multiple testing.

Ethics

The Regional Committee for Medical and Health Research Ethics (REC North) approved the NOWAC Study, the storage of blood samples, and the gene expression analyses in the present study. The Norwegian Data Inspectorate approved the linkages to the Cancer Registry of Norway. Participants received written information about the study and their right to withdraw. Signing the informed consent form, or completing a written questionnaire and donating a blood sample, was regarded as informed consent for cohort enrollment.

Results

Participant characteristics

Mean age at blood sample collection among cases and controls was 56.5 years; mean age at EOC diagnosis among cases was 59.3 years. Cases and controls did not differ significantly with regard to questionnaire variables. Both cases and controls tended toward being overweight, with a mean BMI of approximately 27, and 23% were current smokers. Fewer cases than controls were nullipara, and more cases than controls had 3–4 children (32% vs 24%), but parity distribution did not differ overall ($p = 0.78$). In both groups, approximately 90% were postmenopausal, 20% were current HRT users, and there were no current OC users (Table 1).

Of the 66 women with EOC, 56 (85%) had any metastasis, nine had no metastasis, and one had unknown metastasis status (Table 2). EOC subtype distribution included endometrioid (6%), clear cell (6%), mucinous (4.5%), other/non-specified histologies (15%), and serous EOC

Table 1. Participant characteristics on day of blood sample collection.

Variable	Cases (n = 66)	Controls (n = 66)	P-value ^a
	Mean (SD) or frequency (%)		
Age (years)	56.5 (3.7)	56.5 (3.7)	-
Time to epithelial ovarian cancer diagnosis (years)	2.8	-	
Body mass index (kg/m ²)	26.8 (6.5)	27.0 (4.5)	0.81
Current smoker	15 (23%)	15 (23%)	-
Number of cigarettes yesterday	11.8 (7.9)	10.1 (8.1)	0.57
Number of cigarettes today	1.9	2.9	0.25
Parity			0.78 ^b
0	6 (9%)	7 (11%)	
1–2	37 (56%)	40 (60%)	
3–4	21 (32%)	16 (24%)	
≥5	2 (3%)	3 (5%)	
Menopausal status			
Pre- or perimenopausal	6 (9%)	9 (14%)	-
Postmenopausal	59 (91%)	56 (86%)	0.44
Current hormone replacement therapy use (%)	13 (20%)	14 (21%)	0.47
Current oral contraceptive use ^c	0	0	

^a p-values obtained from a two-sided t-test.

^b p-value comparing the distribution of number of children among cases and controls was obtained from a chi-square test.

^c 38% missing values.

<https://doi.org/10.1371/journal.pone.0256442.t001>

(n = 45), which constituted 68% of all EOC (24 with blood sample collected ≤3 years and 21 collected >3 years before diagnosis) (Table 2). Among serous EOC, the percentage of high, low, and unknown grade was 69%, 4%, and 27%, respectively. Of the high-grade serous EOC, 40% and 60% had blood samples collected ≤3 years and >3 years before diagnosis, respectively. Among those with low- and unknown-grade serous EOC, the corresponding distributions were 50% and 50%, and 60% and 40%, respectively. Compared to controls, cases had larger estimated mean fractions of CD8+ T cells and plasma cells (p = 0.08 and p = 0.07, respectively), and smaller fractions of neutrophils, monocytes, and resting mast cells (p = 0.06, p = 0.08, and p = 0.02, respectively; S1 Table).

Table 2. Distribution of epithelial ovarian cancer (EOC) cases in analytical groups (bold text) of case-control differences in gene expression.

Interval from blood sample collection to diagnosis	≤3 years	>3 Years	Sum
All EOC	35^a	31	66
Metastatic EOC	30	26	56
Non-metastatic EOC	5	5	10 ^b
Serous EOC	24	21	45
Non-serous EOC	11	10	21

^aOne case was diagnosed same month as sample collection and this case-control pair was excluded in the single-gene linear models of samples collected ≤3 years before diagnosis.

^bOne case with unknown metastasis status was categorized as non-metastatic.

<https://doi.org/10.1371/journal.pone.0256442.t002>

Table 3. Probes with the 20 greatest absolute log₂FC values^a among the 100^b lowest p-values in single-gene linear models (case-control) of prospective blood samples from all cases of epithelial ovarian cancer (EOC) and subgroup analyses by clinicopathologic characteristics and interval to diagnosis.

All EOC N = 66	log ₂ FC	Metastatic EOC n = 56	log ₂ FC	Serous EOC n = 45	log ₂ FC	≤3 years n = 34	log ₂ FC	>3 years n = 31	log ₂ FC
Positive log₂FC									
<i>GZMH</i>	0.31	<i>LOC642161</i>	0.16	<i>BTN3A2</i>	0.20	<i>DEFA1B</i>	0.64	<i>LEF1</i>	0.21
<i>SNHG5</i>	0.25	<i>CD2</i>	0.15	<i>LOC642161</i>	0.19	<i>C21orf7</i>	0.25	<i>ETS1</i>	0.20
<i>MIAT</i>	0.15	<i>EEF1G</i>	0.13	<i>CD7</i>	0.16	<i>DEFA4</i>	0.23	<i>EEF1G</i>	0.17
<i>LOC642161</i>	0.14	<i>GIMAP5</i>	0.13	<i>GIMAP5</i>	0.16	<i>MIAT</i>	0.20	<i>GLO1</i>	0.16
<i>CD8A</i>	0.12	<i>CD3E</i>	0.13	<i>RPL8</i>	0.13	<i>MCOLN2</i>	0.19	<i>EEF1A1</i>	0.14
<i>RPL8</i>	0.11	<i>RPL8</i>	0.12	<i>LOC387882</i>	0.10	<i>ASCL2</i>	0.16	<i>EDG1</i>	0.13
<i>LOC728855</i>	0.11	<i>CD8A</i>	0.12	<i>KLHDC4</i>	0.10	<i>DGKQ</i>	0.15	<i>NUP88</i>	0.12
<i>APOBEC3G</i>	0.09	<i>APOBEC3G</i>	0.11	<i>TSEN54</i>	0.09	<i>LOC642161</i>	0.15	<i>C10orf32</i>	0.11
<i>LOC387882</i>	0.08	<i>CPT1B</i>	0.10	<i>HERC1</i>	0.09	<i>MT1X</i>	0.13	<i>CCT8</i>	0.11
<i>RAB11FIP5</i>	0.08	<i>NELF</i>	0.09	<i>SAMD3</i>	0.09	<i>MT1E</i>	0.13	<i>EXOSC8</i>	0.11
Negative log₂FC									
<i>PPT1</i>	-0.13	<i>RHOQ</i>	-0.15	<i>TMEM154</i>	-0.19	<i>C20orf111</i>	-0.14	<i>SCAP</i>	-0.17
<i>NA (AL080095)</i>	-0.13	<i>NA (AL080095)</i>	-0.15	<i>TAOK1</i>	-0.19	<i>LAT2</i>	-0.14	<i>TRPC4AP</i>	-0.18
<i>MPPE1</i>	-0.13	<i>FLJ22662</i>	-0.15	<i>FLJ22662</i>	-0.20	<i>TRIB1</i>	-0.15	<i>ELANE</i>	-0.19
<i>TMEM154</i>	-0.13	<i>MPPE1</i>	-0.17	<i>IGSF6</i>	-0.20	<i>OSBPL8</i>	-0.15	<i>ANXA11</i>	-0.19
<i>TAOK1</i>	-0.13	<i>TAOK1</i>	-0.18	<i>RYBP</i>	-0.21	<i>RRP7A</i>	-0.15	<i>VCL</i>	-0.19
<i>USF1</i>	-0.14	<i>TMEM154</i>	-0.18	<i>CD93</i>	-0.23	<i>CYBRD1</i>	-0.16	<i>TSPAN9</i>	-0.21
<i>LAIR2</i>	-0.16	<i>TMEM154-dupl</i>	-0.18	<i>FCGR3B</i>	-0.25	<i>CXCR5</i>	-0.17	<i>CD93</i>	-0.22
<i>LAIR2-dupl</i>	-0.17	<i>CD93</i>	-0.20	<i>KCTD12</i>	-0.25	<i>FLJ22662</i>	-0.18	<i>TAOK1</i>	-0.22
<i>CD93</i>	-0.17	<i>FCGR3B</i>	-0.21	<i>PI3</i>	-0.40	<i>SKAP2</i>	-0.18	<i>USF1</i>	-0.25
<i>KCTD12</i>	-0.20	<i>KCTD12</i>	-0.24	<i>LOC644936</i>	-0.53	<i>PPT1</i>	-0.21	<i>GP9</i>	-0.38

Bold text indicates probes that occurred among the 10 probes for all EOC as well as another group.

^aUnadjusted p-values for the displayed probes ranged from 0.001 to 0.03, and were lowest in metastatic EOC.

^bThe 100 probes with lowest p-values are listed by p-value in S2–S6 Tables.

<https://doi.org/10.1371/journal.pone.0256442.t003>

Among the 100 probes with the lowest p-values, few log₂FC values exceeded ±0.2 (Table 3; S2 Fig shows the volcano plot for all EOC). The largest absolute log₂FC values observed were for *DEFA1B* (log₂FC = 0.64, p = 0.01) in blood samples collected ≤3 years before diagnosis, and *LOC644936* (log₂FC = -0.53, p = 0.02) in serous EOC. These probes did not occur among the 100 lowest p-values in any other group.

No questionnaire variables were significantly associated with case-control status (Table 1) or with gene expression overall (p > 0.12). The estimated leukocyte fractions found to be associated with case-control status (neutrophils, CD8+ T cells, monocytes, resting mast cells, and plasma cells; S1 Table) were associated with gene expression overall (p = 0.02, 0.04, 1.75e-11, 3.00e-05, 5.00e-06, respectively). Therefore, the adjusted gene expression model included these five leukocyte types and no questionnaire variables. The lists of 100 probes with lowest p-values resulting from the unadjusted and adjusted models of all EOC (S2 and S7 Tables) overlapped by 29 probes.

Gene Ontology enrichment. S8 Table displays GO categories related to biological processes overrepresented among the 100 probes with the lowest p-values in all EOC, metastatic EOC, serous EOC, and in blood samples collected ≤3 years or >3 years before diagnosis. Fig 3 presents the GO categories with the lowest p-values in each group, as well as GO categories that overlapped between the groups. In all EOC, the main enriched categories were “execution phase of apoptosis” and “intrinsic apoptotic signaling pathway in response to oxidative stress”

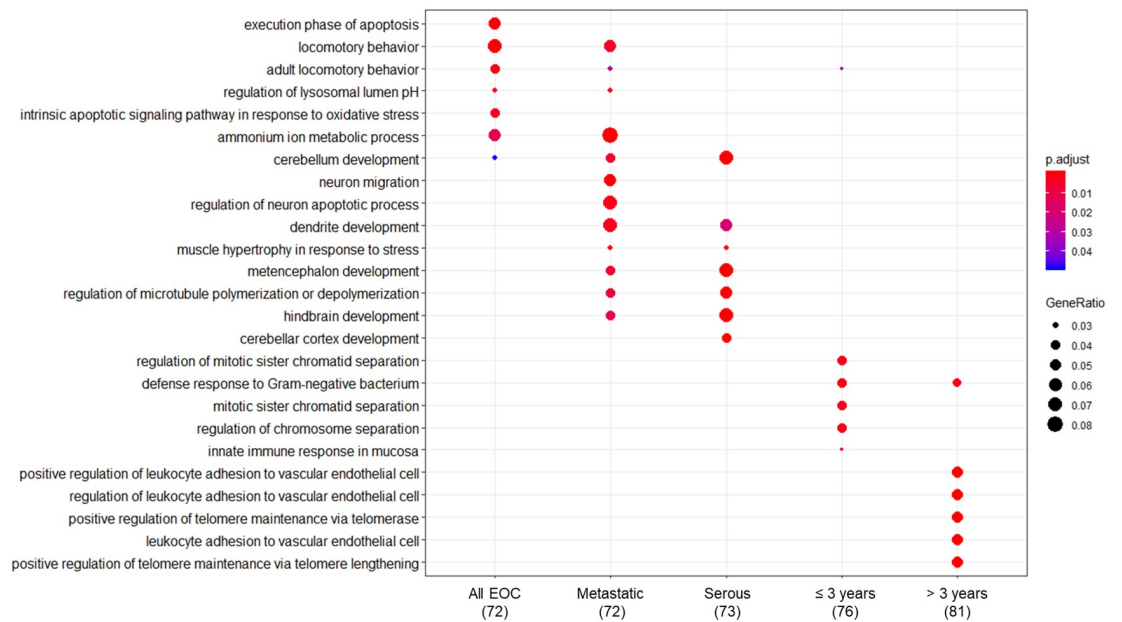


Fig 3. Gene Ontology (GO) enrichment of biological processes among the 100 probes with the lowest p-values in single-gene linear models (case-control) of blood samples from all cases of epithelial ovarian cancer (66 pairs) and according to metastasis status (56 pairs), serous subtype (45 pairs), and interval to diagnosis (≤ 3 years or > 3 years; 34 and 31 pairs, respectively). Numbers below each column indicate the number of probes for which GO categories could be found. The figure presents the five GO categories with lowest p-values. Enriched GO categories ($p < 0.05$) beyond the top five are included in addition if they are among the five most enriched of one of the other investigated groups. [S8 Table](#) presents the complete GO enrichment list.

<https://doi.org/10.1371/journal.pone.0256442.g003>

(contributing genes: *TAOK1*, *STK24*, *RFFL*, *HTRA2*, *DIABLO*), “locomotory behavior” (*CLN6*, *NR4A2*, *PPT1*, *PDE1B*, *HTRA2*), and “regulation of lysosomal lumen pH” (*CLN6*, *PPT1*). With the exception of *PDE1B*, these probes displayed negative \log_2FC values.

In metastatic EOC, the main enriched categories were “ammonium ion metabolic process” (*NR4A2*, *PLA2G7*, *PDE1B*, *PLBD1*, *CHKA*, *CPT1B*), “neuron migration” and “regulation of neuron apoptotic process” (*NR4A2*, *MEF2C*, *PPT1*, *CDK5R1*, *NSMF*), and “dendrite development” (*EZH2*, *MEF2C*, *CDK5R1*, *NSMF*, *CD3E*). Among these transcripts, *PDE1B*, *CD3E*, *EZH2*, and *NSMF* displayed positive \log_2FC values.

In serous EOC, enriched GO categories were “cerebellum development”, “metencephalon development”, “hindbrain development”, “cerebellar cortex development” (*CDK5R1*, *PAK1*, *EZH2*, *SERPINE2*, *HERC1*), and “regulation of microtubule polymerization or depolymerization” (*CDK5R1*, *PAK1*, *FES*, *TAOK1*). The four genes associated with microtubule polymerization displayed negative \log_2FC values, while the remaining \log_2FC values were positive. In blood samples collected ≤ 3 years before diagnosis, the main enriched GO categories were “regulation of mitotic sister chromatid separation” (*PTTG3P*, *CENPE*, *PTTG1*), “defense response to Gram-negative bacterium” (*DEFA1B*, *DEFA4*, *TNFRSF14*), and “innate immune response in mucosa” (*DEFA1B*, *DEFA4*). These probes displayed positive \log_2FC values, and none of the genes were included in the top five enriched categories of the other groups.

Finally, in samples collected > 3 years before diagnosis, the enriched GO categories were “positive regulation of leukocyte adhesion to vascular endothelial cell” (*NFAT5*, *ICAM1*, *ELANE*, *ETS1*) and “positive regulation of telomere maintenance via telomerase and telomere lengthening” (*CCT2*, *CCT8*, *MAPKAPK5*, *HMBOX1*). In the first mentioned GO category, all

probes except *ETS1* displayed negative \log_2FC values, whereas in the latter, all except *HMBOX1* were positive.

Differential expression of genes identified in published functional genomics studies.

Our metastatic EOC group contained expression values for 42 of the 86 genes from relevant publications. S9 Table lists the genes, the \log_2FC values we observed for these probes, and the difference in expression or methylation status in the original studies. From the two gene sets obtained from whole blood gene expression studies, our data contained expression values for two of six genes identified by qPCR [10] and five of six genes previously identified using gene expression microarrays [11]. The lowest p-values we observed from these gene sets were for the probes *CTNNA1* ($\log_2FC = -0.05$, $p = 0.09$) and *NCALD* ($\log_2FC = 0.08$, $p = 0.08$).

Our dataset contained expression values for genes adjacent to more than two-thirds of the methylation sites identified by Teschendorff et al. [5] and Koestler et al. [7]. Among these gene sets, three probes had p-values <0.05 : *LIME1* ($\log_2FC = 0.11$, $p = 0.05$) and *GPR162* ($\log_2FC = -0.17$, $p = 0.04$) from Teschendorff et al. [5], and *STAB1* ($\log_2FC = -0.05$, $p = 0.01$) from Koestler et al. [7]. Our dataset contained expression values for less than half of the genes adjacent to methylation sites identified by Fridley et al. [6], Li et al. [9], and Yang et al. [8]. Among these, the test of *SKAPI* identified by Yang et al. [8] resulted in a p-value <0.05 ($\log_2FC = 0.07$, $p = 0.04$).

Discussion

This nested case-control study of gene expression in whole blood collected up to 7 years prior to EOC diagnosis revealed no statistically significant global or gene-wise associations with EOC case status. The data were high-dimensional, which hampered the statistical power, and the sample size limited the possibilities for analyses according to tumor characteristics or time intervals. Nevertheless, group differences in p-values indicated smaller variation in analyses restricted to metastatic EOC or serous EOC, and greater variation in blood samples collected ≤ 3 years before diagnosis. Compared to controls, cases had larger estimated mean fractions of CD8+ T cells and plasma cells and smaller fractions of neutrophils, monocytes, and resting mast cells. Adjusting for these differences altered the ranking of probes by p-value, but otherwise did not change the results. In targeted gene-wise tests of 42 genes associated with EOC in previous genetic, epigenetic, and transcriptomic studies in blood, four genes were nominally significant among the metastatic cases in the present study.

Case-control differences in gene expression

Neither unsupervised clustering methods, the global test, single-gene linear models, nor gene sets identified statistically significant case-control differences in blood gene expression. With the exception of a few probes, the \log_2FC values obtained in gene-wise linear models were less than ± 0.2 . A \log_2FC value of 0.2 equals a fold change of 1.15, which, if interpreted as an indicator of effect size in epidemiological terms, corresponds to a 15% increase in risk of disease.

As no genes were significantly differentially expressed in this study, the interpretation of single genes was kept to a minimum. The probe with lowest p-value in all EOC, *ENSA*, was also among the 100 probes with the lowest p-values in metastatic EOC and serous EOC, and displayed a larger \log_2FC value in blood samples collected >3 years before diagnosis. *ENSA* encodes α -endosulfine, a cytoplasmic unstructured phosphoprotein with various binding partners depending on cellular context, and regulatory functions depending on its phosphorylation state [37]. Its functions include regulation of cell cycle and platelet activity [38]. In relation to EOC, a small study of serum autoantibodies detected in women with EOC has indicated *ENSA* as a potential autoantigen [39].

Among the probes with highest \log_2FC values in all EOC were four genes (*GZMH*, *APOBEC3G*, *SNHG5*, *MIAT*) that have previously been indicated in studies targeting EOC. In a network analysis of serum proteins, EOC case status was associated with levels of granzyme H (*GZMH*) in blood samples collected >34.5 months prior to diagnosis [40]. A study of tumor transcriptome data associated quantities of the long, non-coding RNAs *SNHG5* and *MIAT* with EOC stage [41], while *APOBEC3G* expression in tumor infiltrating lymphocytes has been associated with EOC survival [42,43]. These transcripts could potentially be of interest in future studies of circulating markers of EOC, but could not be considered as associated with EOC in our whole transcriptome analysis.

Case-control differences by metastasis status, histological subtype, and interval to diagnosis. The majority of the cases in this study were metastatic at diagnosis, and the majority of the metastatic cases were of serous subtype. The lower p-values in these subgroup analyses compared to all EOC indicated less variation in gene expression between blood samples from women with similar tumor characteristics. Previous studies in the NOWAC postgenome cohort that investigated prospective blood samples from women diagnosed with breast cancer [44] and lung cancer [45] found significant case-control differences in gene expression when analyses were restricted to metastatic cancers. It is uncertain whether the lower p-values we observed for metastatic EOC compared to all EOC reflects a similar phenomenon that would have reached statistical significance with a larger sample size.

Our study was based on blood and would detect signals of cancer developing in the ovaries only by association with the composition of the blood transcriptome. Since serous EOC in particular tends to spread while at a low volume [3], early changes in peripheral immune cells could potentially be a more sensitive systemic indicator of malignant disease than substances of tumor origin, which are produced in proportion to tumor mass [9].

The interval from blood sampling to diagnosis in the present study covers the estimated duration of the development of serous EOC from *in-situ* to stage IV metastatic disease [3]. Inferring from the estimations of Brown and Palmer [3], the women in our study who were diagnosed with serous EOC and had blood samples collected ≤ 3 years before their diagnosis likely suffered from some degree of metastasis at the time of sample collection. Assuming a rapid development of the tumor in the final year before diagnosis [3], the higher p-values and larger \log_2FC values we observed in samples collected ≤ 3 years before diagnosis could reflect larger transcriptional variation in this group, possibly as an indicator of disease-associated transcriptional dysregulation. The percentage of probes with positive \log_2FC values was 70% in this group, compared to 50% in other groups except for all EOC adjusted for leukocyte populations, where this percentage was also 70%. This could suggest a general upregulation of gene transcription in samples collected ≤ 3 years before diagnosis, rather than a specific composition of leukocyte types.

In the samples collected >3 years before diagnosis, which could theoretically contain signals of stage I and II serous EOC [3], the case-control differences in gene expression were not as strong. When comparing the 100 probes with the lowest p-values in samples collected ≤ 3 years and >3 years before diagnosis, no overlap in probes was observed. These groups were similar with regard to the distribution of metastatic and serous EOC. Thus, we observed no common transcriptional profile associated with EOC across the postulated time frame for its development. A recent study used mouse models to confirm shifts in systemic immune status during cancer development [46], and it is possible that if our analyses were designed to capture the dynamics of the disease course, we would have been able to identify similar changes associated with EOC. However, due to the small number of samples, we chose not to perform analyses of shorter time intervals.

Gene Ontology enrichment. To explore whether metastasis status, EOC subtype, or time to diagnosis were reflected in biological processes in blood, we compared overrepresented GO categories from the 100 probes with the lowest p-values in single-gene linear models. The overlap of gene lists and shared GO categories (Figs 1 and 2) reflected that all EOC, metastatic EOC, and serous EOC were nested and largely contained the same samples, and that samples collected ≤ 3 years and > 3 years before diagnosis simply represent subdivisions of all EOC.

Among the GO categories indicated in all, metastatic, or serous EOC, locomotory behavior, neuronal migration and central nervous system development have been designated as relevant for the immune system [47]. Migration is a feature of developing neural cells that immune cells share [48]. Overlapping functions of these genes in the immune and neural systems also include the cellular apparatuses related to signaling pathways and cell-to-cell communication [49,50]. Microtubule polymerization and depolymerization, which was enriched in serous EOC, is intrinsic to lymphocyte migration, but also to formation of the immunological synapses necessary for activation of T and B cells [51]. Thus, the main common feature of the overrepresented GO categories for all EOC, metastatic EOC, and serous EOC was their relation to locomotion. If this observation is related to case status, it could suggest that leukocyte migration is affected by EOC.

For blood samples collected ≤ 3 years before diagnosis, “innate immune response in mucosa” and “defense response to Gram-negative bacterium” were among the main enriched GO categories. Interestingly, “defense response to Gram-negative bacterium” was also overrepresented in blood samples collected > 3 years before diagnosis, though neither samples nor probes overlapped. If linked to EOC, the \log_2 FC values were suggestive of initial downregulation of this process, followed by upregulation closer to diagnosis.

In blood samples collected > 3 years before diagnosis, the categories “positive regulation of leukocyte adhesion to vascular endothelial cell” and “positive regulation of telomere maintenance via telomerase and telomere lengthening” were overrepresented. Telomere maintenance is activated during proliferation of activated T and B cells [52]. While this observation is epidemiologically relevant [52], it could be related to the larger proportion of CD8+ T cells in cases overall. Adhesion to endothelial cells is a core mechanism of leukocyte migration, which adds to the above mentioned results for metastatic and serous EOC.

The RNA species investigated in this study included mRNA and polyadenylated long non-coding RNA, and comprised the transcriptome of all circulating immune cells as well as circulating extracellular RNA. Whole blood transcriptomics may thus offer insight into systemic disease processes or enable discovery of circulating markers of disease. Our study design and sample collection were aimed at performing such explorative analyses; however, our study sample was small, and small differences in expression between cases and controls resulted in gene lists that likely included noise. It has been emphasized that GO databases include certain genes that are annotated to many categories [53] and represent current knowledge of genes. Therefore, we have interpreted GO categories with caution.

Estimated leukocyte fractions. The estimated relative sizes of leukocyte populations varied considerably between individuals. On a 10% significance level, EOC cases had slightly larger fractions of CD8+ T cells and plasma cells (adaptive immune system), and smaller fractions of neutrophils, monocytes, and resting mast cells (innate immune system) compared to controls. Adjusting our gene expression models for these leukocyte proportions altered the probes with the lowest p-values, indicating that genes with expression differences according to case-control status were due to differences in these populations.

EOC has been associated with altered proportions of CD8+ T cells, monocytes, and granulocytes (neutrophils, eosinophils, basophils) at diagnosis [7,54], but these studies reported case-control differences opposite to our estimates. Our non-significant observation of higher

proportions of regulatory T cells and M2 macrophages in cases (S1 Table) is more in line with previous studies (summarized in [4]). It is possible that our mean estimates conceal a time-dependent shift during the prediagnostic interval, or that we did not estimate the cell types most relevant for EOC [55].

We estimated relative proportions of 22 leukocyte types. The estimates diverged from the normal physiological range [56] in a manner similar to a divergence observed in other recent studies in the NOWAC postgenome cohort [45,57], which indicates bias. The source might be the deconvolution matrix [58] or upstream laboratory or data processing.

Differential expression of genes identified in published functional genomics studies.

Finally, we used the metastatic EOC group to assess signatures from previous studies of post-diagnostic blood samples from women with EOC. These genes of interest were identified in gene expression studies of patients grouped by tumor characteristics [10,11], or DNA methylation studies of EOC cases and controls [5–9]. Although study designs differed, we could assess how these genes associate with EOC on the transcriptional level in prediagnostic samples. Targeted analyses also let us overcome the problem of multiple testing that arises in explorative analyses.

Gene-wise tests of 42 genes resulted in four probes (*LIME1*, *GPR162*, *STAB1*, *SKAPI*) with p-values <0.05 (S9 Table). We observed the largest \log_2 FC values for *LIME1* and *GPR162* from the study by Teschendorff et al. [5]. *LIME1* (Lck interacting transmembrane adaptor 1; \log_2 FC = 0.11) is expressed in T cells and B cells, where it links T and B cell receptors to downstream signaling pathways via kinases in the Src family [59]. *GPR162* (G Protein-Coupled Receptor 162; \log_2 FC = -0.17) encodes an orphan receptor with adrenaline and noradrenaline as putative ligands [60]. Its mRNA is enriched in neutrophils, monocytes and fallopian tube, but the protein is primarily expressed in the brain [59]. Teschendorff et al. [5] partially attributed the methylation differences they observed to tumor-associated changes in circulating leukocyte composition, and they reported hypermethylation of *LIME1* and *GPR162* in EOC cases. We observed divergent \log_2 FC values for these probes, which, considering the cell type specificity of the transcripts, was in line with our estimated differences in leukocyte populations. However, if the expression difference we observed for *GPR162* is partially attributable to a global change in methylation, this could suggest an altered reception of adrenergic signaling [61–63].

STAB1 (Stabilin 1; \log_2 FC = -0.05) from the study of Koestler et al. [7] encodes a scavenger receptor suggested to mark immunosuppressive monocytes and macrophages, where decreased expression appears to increase T cell antitumor cytotoxicity [64]. *SKAPI* (Src kinase-associated phosphoprotein 1; \log_2 FC = 0.07) from the gene set of Yang et al. [8] encodes a T cell receptor adaptor protein and is a known EOC risk locus with a possible cell-autonomous role in EOC tumorigenesis [65]. Yang et al. [8] reported two methylation sites for this gene in leukocytes: one site was associated with higher *SKAPI* expression and higher EOC risk, and the other with lower *SKAPI* expression and lower EOC risk. Our observation supports a positive association between EOC and levels of *SKAPI* transcripts in blood, though this could simply reflect the proportion of T cells in our study.

In summary, the genes with nominally significant differential expression coded for receptor proteins and for adaptor proteins involved in Src pathways. These genes derived from methylation signatures of EOC predisposition or early disease [5] and methylation-mediated genetic risk [7,8].

Strengths and weaknesses

The main weakness of this study is its sample size, which hampered the power of the statistical analyses and limited the methodological possibilities for modeling continuous relationships

between gene expression and time to diagnosis. We excluded borderline epithelial tumors *a priori*, which further reduced the sample size. These tumors could have been included as non-metastatic EOC, but they represent a pathological entity separate from invasive carcinomas. We did not evaluate potential confounding by past exposure to exogenous hormones. Further, the NOWAC postgenome cohort has not contributed repeat blood samples at different time points during follow-up, a practice which has proven useful in linking proteomic data to EOC [40]. The present study was designed to be explorative and descriptive. Even though any findings might have been useful for biomarker development, the sample size in this study was insufficient to adopt a training, validation and test approach. There were no clear candidate transcripts to pursue in further analyses as potential biomarkers.

Strengths of this study include an epidemiological design aimed at avoiding sampling bias, and blood sample collection during a period that addresses the need for data on circulating molecular markers from women with early-stage EOC. Further, the case-control pairs were matched on age and sample storage time, and we evaluated potential confounding by leukocyte proportions and risk factors.

We chose an analytical approach commonly used in gene expression studies, and which was in line with another whole blood gene expression study related to EOC [11]. The small case-control differences implies that potential signals in the data are subtle against a noisy background; the data are high-dimensional and the results non-significant when adjusted for multiple testing.

Conclusion

This nested case-control study did not reveal statistically significant differences in the peripheral blood transcriptome prior to a diagnosis of EOC. The exploration of transcriptional profiles in blood indicated case-control differences that were small in magnitude and did not reach statistical significance when adjusted for multiple testing. The estimated leukocyte population distributions suggested larger proportions of adaptive immune cell types and smaller proportions of innate immune cell types in cases than in controls, and the functional enrichment suggested lower expression of genes involved in migration. Blood samples collected ≤ 3 years before diagnosis, a larger proportion of which likely represented cases who suffered from advanced EOC, displayed a somewhat larger variation and magnitude in expression, yet we did not observe statistically significant case-control differences in gene expression. Among genes previously linked to ovarian cancer, tests of *LIME1*, *GPR162*, *STAB1*, and *SKAP1* resulted in unadjusted p-values < 0.05 .

The prospective, population-based sampling was a major strength of this study, but the statistical power for explorative transcriptomics was limited. Including a greater number of samples or repeated measurements will allow closer investigation of whether transcript levels change during the course of EOC development.

Supporting information

S1 Fig. No separation of epithelial ovarian cancer cases and controls in hierarchical clustering of gene expression data. Cases shown in orange, controls in cyan. Dendrogram based on \log_2 FC values of the 500 probes with lowest p-values in single-gene linear models of each case-control pair in all EOC. (JPG)

S2 Fig. No separation of epithelial ovarian cancer cases and controls in multidimensional scaling of gene expression data. Cases shown in orange, controls in cyan. Plot based on

\log_2 FC values of the 500 probes with lowest p-values in single-gene linear models of each case-control pair in all EOC.

(PNG)

S3 Fig. Small differences in gene expression among the probes with the lowest p-values.

Few \log_2 FC values exceeded ± 0.2 . Volcano plot of \log_2 FC values and p-values of the 100 probes with lowest p-values in single-gene linear models of all EOC.

(PNG)

S1 Table. Mean estimated fractions of leukocyte populations in blood samples from all cases of epithelial ovarian cancer and controls. Based on deconvolution of gene expression values. P-value from a two-sided t-test of the mean difference.

(XLSX)

S2 Table. The 100 probes with the lowest p-values in single-gene linear models (case-control) of gene expression in blood samples from cases of epithelial ovarian cancer (66 pairs). The presented p-values are not adjusted for multiple testing. All FDR q-values > 0.96 .

(XLSX)

S3 Table. The 100 probes with the lowest p-values in single-gene linear models (case-control) of gene expression in blood samples from cases of metastatic epithelial ovarian cancer (56 pairs). The presented p-values are not adjusted for multiple testing. All FDR q-values > 0.96 .

(XLSX)

S4 Table. The 100 probes with the lowest p-values in single-gene linear models (case-control) of gene expression in blood samples from cases of serous epithelial ovarian cancer (45 pairs). Almost all serous cases were metastatic. The presented p-values are not adjusted for multiple testing. All FDR q-values > 0.96 .

(XLSX)

S5 Table. The 100 probes with the lowest p-values in single-gene linear models (case-control) of gene expression in blood samples collected ≤ 3 years before diagnosis (34 pairs). The presented p-values are not adjusted for multiple testing. All FDR q-values > 0.96 .

(XLSX)

S6 Table. The 100 probes with the lowest p-values in single-gene linear models (case-control) of gene expression in blood samples collected > 3 years before diagnosis (31 pairs). The presented p-values are not adjusted for multiple testing. All FDR q-values > 0.96 .

(XLSX)

S7 Table. The 100 probes with the lowest p-values in single-gene linear models (case-control) of blood samples from all cases of epithelial ovarian cancer (EOC) (66 pairs) in models adjusted for leukocyte populations. Adjusted for estimated fractions of resting mast cells, plasma cells, neutrophils, monocytes, and CD8+ T cells (S1 Table). The presented p-values are not adjusted for multiple testing; all FDR q-values > 0.96 .

(XLSX)

S8 Table. Background data for Fig 2. Enriched Gene Ontology (GO) categories for biological processes among the 100 probes with the lowest p-values in single-gene linear models (case-control) of blood samples from all cases of epithelial ovarian cancer (EOC), metastatic

EOC (56 pairs), serous EOC (45 pairs, almost all were metastatic), and from blood samples collected ≤ 3 years or > 3 years before diagnosis (34 and 31 pairs, respectively).
(XLSX)

S9 Table. Summary of tests of genes identified in published functional genomics studies.

Results from targeted tests of single genes identified in published studies investigating gene expression in peripheral whole blood or DNA methylation in circulating leukocytes from women with epithelial ovarian cancer.
(XLSX)

Acknowledgments

We acknowledge the women who participated in the study. Finances, data collection and data management were administered by Merete Albertsen, Bente Augdal and Knut Hansen. The laboratory analyses were carried out by the Genomics Core Facility (GCF), Norwegian University of Science and Technology (NTNU), Trondheim, Norway.

Disclaimer

Some of the data in this article are from the Cancer Registry of Norway. The Cancer Registry of Norway is not responsible for the analysis or interpretation of the data presented.

Author Contributions

Conceptualization: Mie Jareid, Eiliv Lund.

Data curation: Therese Haugdahl Nøst.

Formal analysis: Therese Haugdahl Nøst.

Funding acquisition: Eiliv Lund.

Investigation: Mie Jareid, Eiliv Lund.

Methodology: Marit Holden.

Resources: Eiliv Lund.

Supervision: Lill-Tove Rasmussen Busund, Therese Haugdahl Nøst.

Writing – original draft: Mie Jareid, Therese Haugdahl Nøst.

Writing – review & editing: Mie Jareid, Igor Snapkov, Marit Holden, Lill-Tove Rasmussen Busund, Eiliv Lund, Therese Haugdahl Nøst.

References

1. Cancer Registry of Norway. Cancer in Norway 2017—Cancer incidence, mortality, survival and prevalence in Norway. Oslo: Cancer Registry of Norway, 2018.
2. Terry KL, Missmer SA. Epidemiology of Ovarian and Endometrial Cancers. In: Loda M, Mucci LA, Mittelstadt ML, Van Hemelrijck M, Cotter MB, editors. Pathology and Epidemiology of Cancer. Cham: Springer International Publishing; 2017. p. 233–46.
3. Brown PO, Palmer C. The preclinical natural history of serous ovarian cancer: defining the target for early detection. *PLoS medicine*. 2009; 6(7):e1000114. Epub 2009/07/29. <https://doi.org/10.1371/journal.pmed.1000114> PMID: 19636370.
4. Coosemans A, Baert T, Ceusters J, Busschaert P, Landolfo C, Verschuere T, et al. Myeloid-derived suppressor cells at diagnosis may discriminate between benign and malignant ovarian tumors. *International journal of gynecological cancer: official journal of the International Gynecological Cancer Society*. 2019; 29(9):1381–8. Epub 2019/11/07. <https://doi.org/10.1136/ijgc-2019-000521> PMID: 31685557.

5. Teschendorff AE, Menon U, Gentry-Maharaj A, Ramus SJ, Gayther SA, Apostolidou S, et al. An epigenetic signature in peripheral blood predicts active ovarian cancer. *PLoS one*. 2009; 4(12):e8274. Epub 2009/12/19. <https://doi.org/10.1371/journal.pone.0008274> PMID: 20019873.
6. Fridley BL, Armasu SM, Cicek MS, Larson MC, Wang C, Winham SJ, et al. Methylation of leukocyte DNA and ovarian cancer: relationships with disease status and outcome. *BMC Medical Genomics*. 2014; 7(1):21. <https://doi.org/10.1186/1755-8794-7-21> PMID: 24774302
7. Koestler DC, Chalise P, Cicek MS, Cunningham JM, Armasu S, Larson MC, et al. Integrative genomic analysis identifies epigenetic marks that mediate genetic risk for epithelial ovarian cancer. *BMC medical genomics*. 2014; 7:8. Epub 2014/02/01. <https://doi.org/10.1186/1755-8794-7-8> PMID: 24479488.
8. Yang Y, Wu L, Shu X, Lu Y, Shu X-o, Cai Q, et al. Genetic data from nearly 63,000 women of European descent predicts DNA methylation biomarkers and epithelial ovarian cancer risk. *Cancer Research*. 2018; canres.2726.018. <https://doi.org/10.1158/0008-5472.CAN-18-2726> PMID: 30559148
9. Li L, Zheng H, Huang Y, Huang C, Zhang S, Tian J, et al. DNA methylation signatures and coagulation factors in the peripheral blood leucocytes of epithelial ovarian cancer. *Carcinogenesis*. 2017; 38(8):797–805. Epub 2017/06/24. <https://doi.org/10.1093/carcin/bgx057> PMID: 28637314.
10. Isaksson HS, Sorbe B, Nilsson TK. Whole blood RNA expression profiles in ovarian cancer patients with or without residual tumors after primary cytoreductive surgery. *Oncology reports*. 2012; 27(5):1331–5. Epub 2012/02/11. <https://doi.org/10.3892/or.2012.1680> PMID: 22322362.
11. Isaksson HS, Sorbe B, Nilsson TK. Whole genome expression profiling of blood cells in ovarian cancer patients -prognostic impact of the CYP1B1, MTSS1, NCALD, and NOP14. *Oncotarget*. 2014; 5(12):4040–9. Epub 2014/06/26. <https://doi.org/10.18632/oncotarget.1938> PMID: 24961659.
12. Pils D, Tong D, Hager G, Obermayr E, Aust S, Heinze G, et al. A combined blood based gene expression and plasma protein abundance signature for diagnosis of epithelial ovarian cancer—a study of the OVCAD consortium. *BMC Cancer*. 2013; 13(1):178. <https://doi.org/10.1186/1471-2407-13-178> PMID: 23551967
13. Mok CS KJ, Skates SJ, Schorge JO, Cramer DW, Lu KH, Liew CC. Use of Blood-based mRNA profiling to Identify Biomarkers for Ovarian Cancer Screening. *Gynecology & obstetrics (Sunnyvale, Calif)*. 2017; 7(6). <https://doi.org/10.4172/2161-0932.1000443>
14. Dumeaux V, Borresen-Dale AL, Frantzen JO, Kumle M, Kristensen VN, Lund E. Gene expression analyses in breast cancer epidemiology: the Norwegian Women and Cancer postgenome cohort study. *Breast cancer research: BCR*. 2008; 10(1):R13. Epub 2008/02/15. <https://doi.org/10.1186/bcr1859> PMID: 18271962.
15. Lund E, Dumeaux V, Braaten T, Hjartaker A, Engeset D, Skeie G, et al. Cohort profile: The Norwegian Women and Cancer Study—NOWAC—Kvinner og kreft. *International journal of epidemiology*. 2008; 37(1):36–41. Epub 2007/07/24. <https://doi.org/10.1093/ije/dym137> PMID: 17644530.
16. Dumeaux V, Børresen-Dale A-L, Frantzen J-O, Kumle M, Kristensen VN, Lund E. Gene expression analyses in breast cancer epidemiology: the Norwegian Women and Cancer postgenome cohort study. *Breast Cancer Res*. 2008; 10(1):R13. <https://doi.org/10.1186/bcr1859> PMID: 18271962.
17. Larsen IK, Småstuen M, Johannesen TB, Langmark F, Parkin DM, Bray F, et al. Data quality at the Cancer Registry of Norway: An overview of comparability, completeness, validity and timeliness. *European Journal of Cancer*. 2009; 45(7):1218–31. <https://doi.org/10.1016/j.ejca.2008.10.037> PMID: 19091545
18. Wild C, Weiderpass E, Stewart B. World Cancer Report: Cancer Research for Cancer Prevention. Lyon, France: International Agency for Research on Cancer. 2020.
19. de Luis DA, Almansa R, Aller R, Izaola O, Romero E. Gene expression analysis identify a metabolic and cell function alterations as a hallmark of obesity without metabolic syndrome in peripheral blood, a pilot study. *Clinical Nutrition*. 2018; 37(4):1348–53. <https://doi.org/10.1016/j.clnu.2017.06.006> PMID: 28633944
20. Huan T, Joehanes R, Schurmann C, Schramm K, Pilling LC, Peters MJ, et al. A whole-blood transcriptome meta-analysis identifies gene expression signatures of cigarette smoking. *Human molecular genetics*. 2016; 25(21):4611–23. Epub 2017/02/06. <https://doi.org/10.1093/hmg/ddw288> PMID: 28158590.
21. Lund E, Nakamura A, Snapkov I, Thalabard JC, Olsen KS, Holden L, et al. Each pregnancy linearly changes immune gene expression in the blood of healthy women compared with breast cancer patients. *Clinical epidemiology*. 2018; 10:931–40. Epub 2018/08/21. <https://doi.org/10.2147/CLEP.S163208> PMID: 30123005.
22. Levine ME, Lu AT, Chen BH, Hernandez DG, Singleton AB, Ferrucci L, et al. Menopause accelerates biological aging. *Proc Natl Acad Sci U S A*. 2016; 113(33):9327–32. Epub 2016/07/25. <https://doi.org/10.1073/pnas.1604558113> PMID: 27457926.
23. Dahm AEA, Eilertsen AL, Goeman J, Olstad OK, Øvstebø R, Kierulf P, et al. A microarray study on the effect of four hormone therapy regimens on gene transcription in whole blood from healthy

- postmenopausal women. *Thrombosis Research*. 2012; 130(1):45–51. <https://doi.org/10.1016/j.thromres.2011.12.009> PMID: 22217510
24. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic acids research*. 2015; 43(7):e47. Epub 2015/01/22. <https://doi.org/10.1093/nar/gkv007> PMID: 25605792.
 25. Du P, Kibbe WA, Lin SM. lumi: a pipeline for processing Illumina microarray. *Bioinformatics*. 2008; 24(13):1547–8. Epub 2008/05/10. <https://doi.org/10.1093/bioinformatics/btn224> PMID: 18467348.
 26. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, et al. Robust enumeration of cell subsets from tissue expression profiles. *Nature Methods*. 2015; 12:453. <https://doi.org/10.1038/nmeth.3337> PMID: 25822800
 27. Goeman JJ, Geer SAVd, Kort Fd, Houwelingen HCv. A global test for groups of genes: testing association with a clinical outcome. *Bioinformatics*. 2004; 20. <https://doi.org/10.1093/bioinformatics/btg382> PMID: 14693814
 28. Wu D, Lim E, Vaillant F, Asselin-Labat ML, Visvader JE, Smyth GK. ROAST: rotation gene set tests for complex microarray experiments. *Bioinformatics (Oxford, England)*. 2010; 26. <https://doi.org/10.1093/bioinformatics/btq401> PMID: 20610611
 29. The Gene Ontology Consortium. The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Research*. 2018; 47(D1):D330–D8. <https://doi.org/10.1093/nar/gky1055> PMID: 30395331
 30. Godec J, Tan Y, Liberzon A, Tamayo P, Bhattacharya S, Butte Atul J, et al. Compendium of Immune Signatures Identifies Conserved and Species-Specific Biology in Response to Inflammation. *Immunity*. 2016; 44(1):194–206. <https://doi.org/10.1016/j.immuni.2015.12.006> PMID: 26795250
 31. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic acids research*. 2000; 28(1):27–30. Epub 1999/12/11. <https://doi.org/10.1093/nar/28.1.27> PMID: 10592173.
 32. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences*. 2005; 102(43):15545–50. <https://doi.org/10.1073/pnas.0506580102> PMID: 16199517
 33. Stangroom J. Social Science Statistics 2019 [3. January 2020]. Chi-Square Test Calculator]. <https://www.socscistatistics.com/tests/chisquare2/Default2.aspx>.
 34. Yu G, Wang L-G, Han Y, He Q-Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics: a journal of integrative biology*. 2012; 16(5):284–7. <https://doi.org/10.1089/omi.2011.0118> PMID: 22455463.
 35. The Gene Ontology Consortium. Expansion of the Gene Ontology knowledgebase and resources. *Nucleic acids research*. 2017; 45(D1):D331–d8. Epub 2016/12/03. <https://doi.org/10.1093/nar/gkw1108> PMID: 27899567.
 36. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature genetics*. 2000; 25(1):25–9. Epub 2000/05/10. <https://doi.org/10.1038/75556> PMID: 10802651.
 37. Boettcher JM, Hartman KL, Lador DT, Qi Z, Woods WS, George JM, et al. Membrane-induced folding of the cAMP-regulated phosphoprotein endosulfine- α . *Biochemistry*. 2008; 47(47):12357–64. Epub 2008/11/01. <https://doi.org/10.1021/bi801450t> PMID: 18973346.
 38. Kumm EJ, Pagel O, Gambaryan S, Walter U, Zahedi RP, Smolenski A, et al. The Cell Cycle Checkpoint System MAST(L)-ENSA/ARPP19-PP2A is Targeted by cAMP/PKA and cGMP/PKG in Anucleate Human Platelets. *Cells*. 2020; 9(2). Epub 2020/02/23. <https://doi.org/10.3390/cells9020472> PMID: 32085646.
 39. Murphy M, O'Connell D, O'Kane S, O'Brien J, O'Toole S, Martin C, et al. Epitope presentation is an important determinant of the utility of antigens identified from protein arrays in the development of auto-antibody diagnostic assays. *J Proteomics*. 2012; 75(15):4668–75. <https://doi.org/10.1016/j.jprot.2012.02.031> PMID: 22415278.
 40. Whitwell HJ, Blyuss O, Menon U, Timms JF, Zaikin A. Parenclitic networks for predicting ovarian cancer. *Oncotarget*. 2018; 9(32):22717–26. Epub 2018/06/02. <https://doi.org/10.18632/oncotarget.25216> PMID: 29854310.
 41. Zhao Q, Fan C. A novel risk score system for assessment of ovarian cancer based on co-expression network analysis and expression level of five lncRNAs. *BMC Med Genet*. 2019; 20(1):103-. <https://doi.org/10.1186/s12881-019-0832-9> PMID: 31182053.
 42. Svoboda M, Meshcheryakova A, Heinze G, Jaritz M, Pils D, Castillo-Tong DC, et al. AID/APOBEC-network reconstruction identifies pathways associated with survival in ovarian cancer. *BMC Genomics*. 2016; 17(1):643. <https://doi.org/10.1186/s12864-016-3001-y> PMID: 27527602

43. Leonard B, Starrett GJ, Maurer MJ, Oberg AL, Van Bockstal M, Van Dorpe J, et al. APOBEC3G Expression Correlates with T-Cell Infiltration and Improved Clinical Outcomes in High-grade Serous Ovarian Carcinoma. *Clinical cancer research: an official journal of the American Association for Cancer Research*. 2016; 22(18):4746–55. Epub 2016/03/27. <https://doi.org/10.1158/1078-0432.Ccr-15-2910> PMID: 27016308.
44. Lund E, Holden L, Bøvelstad H, Plancade S, Mode N, Günther C-C, et al. A new statistical method for curve group analysis of longitudinal gene expression data illustrated for breast cancer in the NOWAC postgenome cohort as a proof of principle. *BMC Medical Research Methodology*. 2016; 16(1):28. <https://doi.org/10.1186/s12874-016-0129-z> PMID: 26944545
45. Nøst TH, Holden M, Dønnem T, Bøvelstad H, Rylander C, Lund E, et al. Transcriptomic signals in blood prior to lung cancer focusing on time to diagnosis and metastasis. *Scientific Reports*. 2021; 11(1):7406. <https://doi.org/10.1038/s41598-021-86879-8> PMID: 33795786
46. Allen BM, Hiam KJ, Burnett CE, Venida A, DeBarge R, Tenvooren I, et al. Systemic dysfunction and plasticity of the immune macroenvironment in cancer models. *Nature medicine*. 2020; 26(7):1125–34. Epub 2020/05/27. <https://doi.org/10.1038/s41591-020-0892-6> PMID: 32451499.
47. Geifman N, Monsonego A, Rubin E. The Neural/Immune Gene Ontology: clipping the Gene Ontology for neurological and immunological systems. *BMC bioinformatics*. 2010; 11:458-. <https://doi.org/10.1186/1471-2105-11-458> PMID: 20831831.
48. Meehan TF, Vasilevsky NA, Mungall CJ, Dougall DS, Haendel MA, Blake JA, et al. Ontology based molecular signatures for immune cell types via gene expression analysis. *BMC Bioinformatics*. 2013; 14(1):263. <https://doi.org/10.1186/1471-2105-14-263> PMID: 24004649
49. Safe S, Jin UH, Morpurgo B, Abudayyeh A, Singh M, Tjalkens RB. Nuclear receptor 4A (NR4A) family—orphans no more. *The Journal of steroid biochemistry and molecular biology*. 2016; 157:48–60. Epub 2015/04/29. <https://doi.org/10.1016/j.jsbmb.2015.04.016> PMID: 25917081.
50. Durocher M, Ander BP, Jickling G, Hamade F, Hull H, Knepp B, et al. Inflammatory, regulatory, and autophagy co-expression modules and hub genes underlie the peripheral immune response to human intracerebral hemorrhage. *Journal of Neuroinflammation*. 2019; 16(1):56. <https://doi.org/10.1186/s12974-019-1433-4> PMID: 30836997
51. Ilan-Ber T, Ilan Y. The role of microtubules in the immune system and as potential targets for gut-based immunotherapy. *Molecular immunology*. 2019; 111:73–82. Epub 2019/04/30. <https://doi.org/10.1016/j.molimm.2019.04.014> PMID: 31035111.
52. de Punder K, Heim C, Przesdzing I, Wadhwa PD, Entringer S. Characterization in humans of *in vitro* leucocyte maximal telomerase activity capacity and association with stress. *Philosophical Transactions of the Royal Society B: Biological Sciences*. 2018; 373(1741):20160441. <https://doi.org/10.1098/rstb.2016.0441> PMID: 29335365
53. Ballouz S, Pavlidis P, Gillis J. Using predictive specificity to determine when gene set analysis is biologically meaningful. *Nucleic acids research*. 2017; 45(4):e20–e. <https://doi.org/10.1093/nar/gkw957> PMID: 28204549.
54. Xiang J, Zhou L, Li X, Bao W, Chen T, Xi X, et al. Preoperative Monocyte-to-Lymphocyte Ratio in Peripheral Blood Predicts Stages, Metastasis, and Histological Grades in Patients with Ovarian Cancer. *Translational Oncology*. 2017; 10(1):33–9. <https://doi.org/10.1016/j.tranon.2016.10.006> PMID: 27888711
55. Stenzel AE, Abrams SI, Moysich KB. A Call for Epidemiological Research on Myeloid-Derived Suppressor Cells in Ovarian Cancer: A Review of the Existing Immunological Evidence and Suggestions for Moving Forward. *Frontiers in immunology*. 2019; 10:1608. Epub 2019/07/30. <https://doi.org/10.3389/fimmu.2019.01608> PMID: 31354741.
56. U. S. National Institutes of Health—National Cancer Institute. SEER Training Modules, Normal Blood Values [cited 2021 Jun 18]. <https://training.seer.cancer.gov/abstracting/procedures/clinical/hematologic/blood.html>.
57. Baiju N, Sandanger TM, Sæstrom P, Nøst TH. Gene expression in blood reflects smoking exposure among cancer-free women in the Norwegian Women and Cancer (NOWAC) postgenome cohort. *Scientific Reports*. 2021; 11(1):680. <https://doi.org/10.1038/s41598-020-80158-8> PMID: 33436844
58. Avila Cobos F, Alquicira-Hernandez J, Powell JE, Mestdagh P, De Preter K. Benchmarking of cell type deconvolution pipelines for transcriptomics data. *Nat Commun*. 2020; 11(1):5650-. <https://doi.org/10.1038/s41467-020-19015-1> PMID: 33159064.
59. proteinatlas.org. The Human Protein Atlas [2020-08-28]. <https://www.proteinatlas.org/>.
60. Kakarala KK, Jamil K. Sequence-structure based phylogeny of GPCR Class A Rhodopsin receptors. *Molecular phylogenetics and evolution*. 2014; 74:66–96. Epub 2014/02/08. <https://doi.org/10.1016/j.ympev.2014.01.022> PMID: 24503482.

61. Predescu DV, Crețoiu SM, Crețoiu D, Pavelescu LA, Suci N, Radu BM, et al. G Protein-Coupled Receptors (GPCRs)-Mediated Calcium Signaling in Ovarian Cancer: Focus on GPCRs activated by Neurotransmitters and Inflammation-Associated Molecules. *International journal of molecular sciences*. 2019; 20(22). Epub 2019/11/11. <https://doi.org/10.3390/ijms20225568> PMID: 31703453.
62. Roberts AL, Huang T, Koenen KC, Kim Y, Kubzansky LD, Tworoger SS. Posttraumatic Stress Disorder Is Associated with Increased Risk of Ovarian Cancer: A Prospective and Retrospective Longitudinal Cohort Study. *Cancer research*. 2019; 79(19):5113–20. Epub 2019/09/07. <https://doi.org/10.1158/0008-5472.CAN-19-1222> PMID: 31488422.
63. Su S, Hao G, Youssef NA, Stefanek M, Treiber F, Wang X, et al. DNA methylation as a mediator of the association between stress and depression: A genome-wide DNA methylation study. *The FASEB Journal* 2019. p. 593.14–.14.
64. Hollmén M, Figueiredo CR, Jalkanen S. New tools to prevent cancer growth and spread: a ‘Clever’ approach. *British Journal of Cancer*. 2020; 123(4):501–9. <https://doi.org/10.1038/s41416-020-0953-0> PMID: 32595212
65. Marcotte R, Brown KR, Suarez F, Sayad A, Karamboulas K, Krzyzanowski PM, et al. Essential gene profiles in breast, pancreatic, and ovarian cancer cells. *Cancer discovery*. 2012; 2(2):172–89. Epub 2012/05/16. <https://doi.org/10.1158/2159-8290.CD-11-0224> PMID: 22585861.