



OPEN

Identification of serum miR-1246 and miR-150-5p as novel diagnostic biomarkers for high-grade serous ovarian cancer

Magdalena Niemira^{1✉}, Anna Erol¹, Agnieszka Bielska¹, Anna Zeller¹, Anna Skwarska², Karolina Chwiałkowska³, Mariusz Kuzmicki⁴, Jacek Szamatowicz⁴, Joanna Reszec⁵, Pawel Knapp⁶, Marcin Moniuszko⁷ & Adam Kretowski¹

Epithelial ovarian cancer (EOC) is one of the leading cancers in women, with high-grade serous ovarian cancer (HGSOC) being the most common and lethal subtype of this disease. A vast majority of HGSOC are diagnosed at the late stage of the disease when the treatment and total recovery chances are low. Thus, there is an urgent need for novel, more sensitive and specific methods for early and routine HGSOC clinical diagnosis. In this study, we performed miRNA expression profiling using the NanoString miRNA assay in 34 serum samples from patients with HGSOC and 36 healthy women. We identified 13 miRNAs that were differentially expressed (DE). For additional exploration of expression patterns correlated with HGSOC, we performed weighted gene co-expression network analysis (WGCNA). As a result, we showed that the module most correlated with tumour size, nodule and metastasis contained 8 DE miRNAs. The panel including miR-1246 and miR-150-5p was identified as a signature that could discriminate HGSOC patients with AUCs of 0.98 and 1 for the training and test sets, respectively. Furthermore, the above two-miRNA panel had an AUC = 0.946 in the verification cohorts of RT-qPCR data and an AUC = 0.895 using external data from the GEO public database. Thus, the model we developed has the potential to markedly improve the diagnosis of ovarian cancer.

Epithelial ovarian cancer (EOC) is diagnosed in an estimated 314,000 women, accounting for 210,000 deaths worldwide annually (GLOBOCAN 2020; <https://doi.org/10.3322/caac.21660>). In 2022, in the United States, over 12,000 women died because of ovarian cancer, including over 9000 at postmenopausal age¹. High-grade serous ovarian cancer (HGSOC) is EOC's most common histological subtype, with a five-year survival rate below 30%. While the disease limited only to the ovaries (stage I) can be cured in up to 90% of patients, most cases are diagnosed at a late stage (stage III or IV)². The routine diagnostic procedures, including transvaginal ultrasonography and CA125 antigen measurement, are unsuitable for early diagnosis^{3,4}. Several protein-based biomarkers, such as CA125 and HE4, are currently being used for FDA-approved diagnosis. The most studied marker, CA125, is only elevated by 50–60% of early-stage EOC or asymptomatic patients. Recent findings suggest that CA125 allows the diagnosis of EOC with 88.7% sensitivity and 74.7% specificity^{5–7}. Combined detection of CA125 and human epididymis secretory protein E4 (HE4) has been shown to improve EOC screening^{8–10}. For example, the Risk of Ovarian Malignancy Algorithm (ROMA) with CA125 and HE4 as biomarkers, showed 100% sensitivity and 74.2% specificity in preoperative assessment^{11,12}. Rein et al.¹³ showed that Ova1 panel testing upregulated CA125-II and B2M and downregulated ApoA1, transthyretin, and transferrin, allowing for EOC diagnosis in 96% of cases with 96% sensitivity and 35% specificity. Given the lack of usual symptoms of EOC and the limitations in early diagnosis identifying more sensitive and specific biomarkers is urgently required.

¹Clinical Research Centre, Medical University of Białystok, Białystok, Poland. ²Cancer Center, Department of Oncology, Albert Einstein College of Medicine, Bronx, NY, USA. ³Centre for Bioinformatics and Data Analysis, Medical University of Białystok, Białystok, Poland. ⁴Department of Gynecology and Gynecological Oncology, Medical University of Białystok, Białystok, Poland. ⁵Department of Medical Pathomorphology, Medical University of Białystok, Białystok, Poland. ⁶University Oncology Centre, University Clinical Hospital in Białystok, Białystok, Poland. ⁷Department of Regenerative Medicine and Immune Regulation, Medical University of Białystok, Białystok, Poland. ✉email: magdalena.niemira@umb.edu.pl

Several groups recently showed the pivotal role of microRNAs (miRNAs) as potential biomarkers for various cancers. Circulating miRNA has many features to fulfil the role of a useful biomarker. They are highly stable and resistant to RNase digestion, high temperatures, extreme pH, extended storage, and multiple freeze–thaw cycles¹⁴. MiRNAs are short (~22 nucleotides), non-coding, single-stranded RNA molecules that play a crucial role in regulating gene expression by interacting with the 3'-UTR region of mRNAs^{15,16}. Since a single miRNA can interact with many mRNAs, miRNAs regulate several cellular processes, such as differentiation, development, proliferation, metabolism, tumorigenesis, apoptosis and angiogenesis^{17–19}. In addition, given the nature of miRNA-regulated genes, these molecules can act as oncogenes (oncomiRs) and tumour suppressors²⁰. However, the current literature often depicts the contradictory role of particular miRNAs in cancer development. It has been hypothesised that such discrepancies may result from the ability of miRNAs to simultaneously affect the expression of both tumour-suppressive and oncogenic mRNAs. For instance, miR-186 may serve as an oncomiR or a tumour suppressor miRNA. Its dual role is related to enhancing or inhibiting proliferation, metastasis, invasion, and apoptosis in cancers²¹. Similarly, while miR-155 was initially identified as an oncomiR in pancreatic cancer²², in ovarian and gastric cancers, its expression is inhibited, indicating that miR-155 may also act as a tumour suppressor²³.

Studies on epithelial ovarian cancer have shown dysregulation of several miRNAs, such as miR-200a, miR-200b, miR-200c and miR-199a in cancer tissue²⁴, and circulating miRNAs in blood, including miR-130-3p, miR-143-3p, miR-145, miR-200c, miR-205-5p, miR-26a-5p, miR-328-3p, miR-374a-5p, and miR-766-3p^{25,26}. Importantly, while numerous studies have already demonstrated the diagnostic or the prognostic potential of circulating miRNAs in ovarian cancer^{27–29}, most of these analyses were performed bulking different subtypes of ovarian cancer regardless of histology. It is worth underlining that different histological subtypes may contribute to contradictory alterations of miRNAs in specific cancers. For example, Schmid et al.³⁰ analysed miR-34a expression in EOC and found that miR-34a was significantly downregulated in serous, endometrioid and mucinous histological subtypes, while no significant alteration in clear cell ovarian cancer was found in comparison with healthy ovary tissues.

Despite many reports, no decisive circulating miRNA signature has been proposed to distinguish HGSOc patients from healthy individuals unambiguously. Herein, we propose that simultaneous identification of miR-1246 and miR-150-5p in serum can be used for the non-invasive diagnosis of HGSOc patients. We also investigated the possible involvement of circulating miRNAs in HGSOc development by analysing the biological importance of miRNA targets and the functional enrichment profile of their target gene sets.

Results

Identification and validation of differentially expressed miRNAs in serum samples of HGSOc patients

Altogether, 13 out of 798 unique miRNAs showed significant differences in counts between HGSOc and healthy donor serum samples (Table 1) with Fold Change (FC) >|1.5| and False Discovery Rates (FDRs) <0.05 (Fig. 1A). Table 2 summarises FCs and FDRs for all thirteen differentially expressed (DE) miRNAs. Five miRNAs were upregulated: miR-1246, miR-4454 + miR-7975 (the mature sequence of miR-7975 differs only by one base from miR-4454), miR-630, and miR-4516, whereas eight miRNAs were downregulated: miR-144-3p, miR-142-3p, miR-150-5p, miR-15a-5p, miR-15b-5p, miR-126-3p, miR-191-5p, and miR-106b-5p. To validate the diagnostic usefulness of the identified miRNAs, we performed RT-qPCR analysis. All DE miRNAs were detectable by quantitative PCR, and the expression data were comparable to those generated using the NanoString nCounter System (Fig. 1B). Next, we used Pearson tests to calculate the correlation coefficient between the expression of miRNAs and the following patients' characteristics: body mass index (BMI), and CA125 serum level. We observed a positive correlation between CA125 level and miR-1246 as well as miR-4454 + miR-7975 expression with correlation coefficients of 0.670 and 0.526, respectively (P value = 8.8×10^{-04} and P value = 1.41×10^{-02}). In addition, there was also a group of miRNAs, which was correlated with each other, including miR-106b-5p, miR-126-3p, miR-142-3p, miR-144-3p, miR-150-5p, miR-15a-5p, and miR-15b-5p (Fig. 3C).

Characteristics	Cases for NanoString analysis						Cases for RT-qPCR validation					
	Training set			Test set			Training set			Test set		
	Cases (n=28)	Controls (n=21)	P val	Cases (n=8)	Controls (n=13)	P val	Cases (n=28)	Controls (n=34)	P val	Cases (n=14)	Controls (n=12)	P val
Age (mean ± SD)	61.6 ± 14	58.1 ± 6	0.2	59.2 ± 6	61.0 ± 4	0.7	61.3 ± 13	59.3 ± 7	0.3	57.7 ± 13	57.8 ± 7	0.9
BMI (mean ± SD)	29.2 ± 6	27.2 ± 3	0.2	27.9 ± 4	28.7 ± 3	0.4	28.8 ± 5	28.0 ± 3	0.9	28.9 ± 6	26.4 ± 2	0.2
FIGO stages, n (%)												
FIGO I	2 (7.1)	–		1 (12.5)	–		3 (10.8)	–		0 (0.0)	–	
FIGO II	0 (0.0)	–		1 (12.5)	–		0 (0.0)	–		1 (7.1)	–	
FIGO III	15 (53.6)	–		3 (37.5)	–		14 (50.0)	–		6 (42.9)	–	
FIGO IV	10 (35.7)	–		3 (37.5)	–		9 (32.1)	–		5 (35.7)	–	
N/A	1 (3.6)	–		0 (0.0)	–		2 (7.1)	–		2 (14.3)	–	

Table 1. Participant characteristics of the NanoString and RT-qPCR model set.

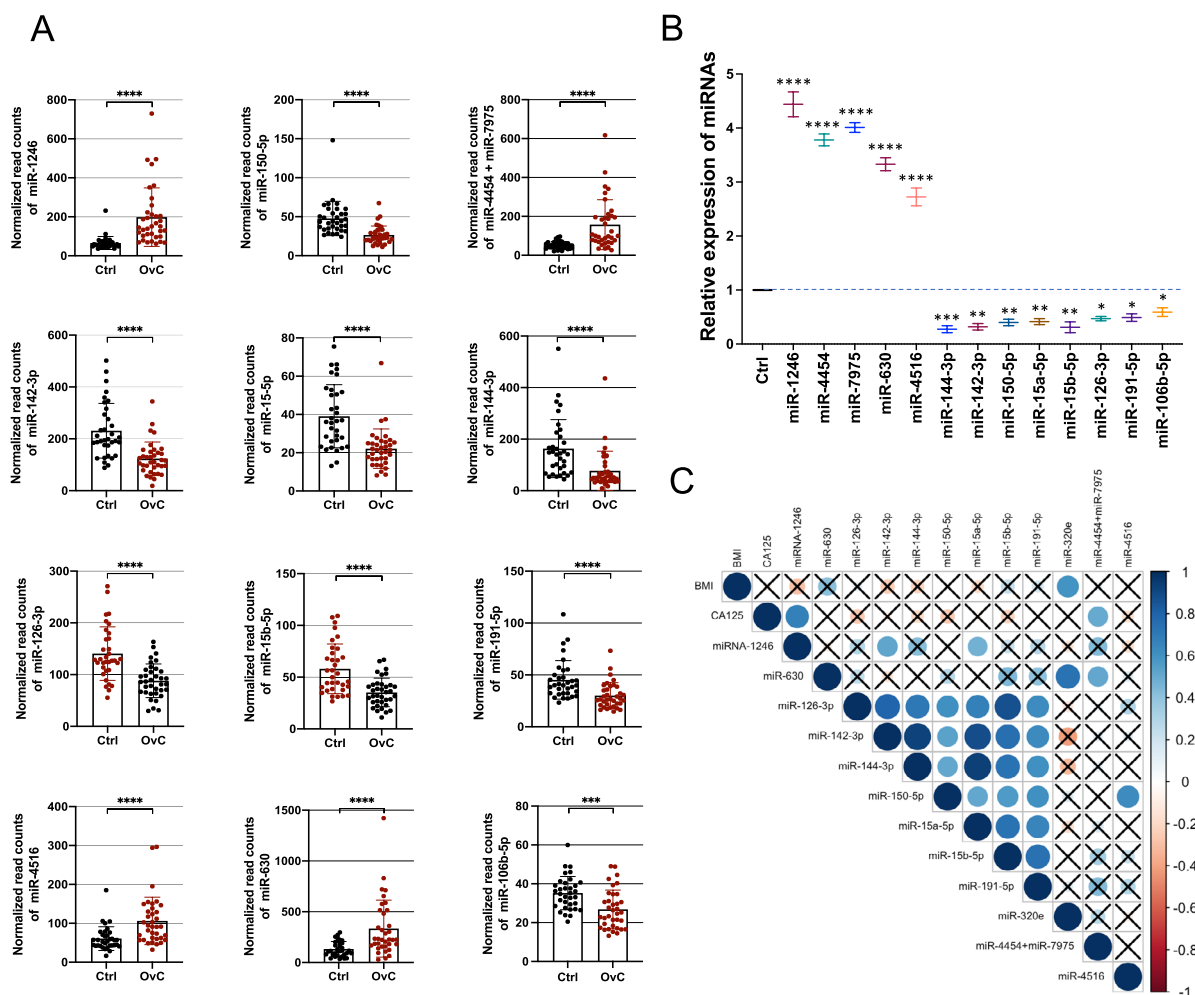


Figure 1. Differential miRNAs expression in the serum of HGSOE patients and healthy control. **(A)** The scatter plot with the bar shows the expression levels of circulating miRNAs in serum in HGSOE compared to healthy control from the NanoString platform. Differences in the expression levels in miRNAs between patients and controls were compared using the Mann–Whitney test; **(B)** RT–qPCR validation of 13 miRNAs selected from the NanoString platform. Each bar represents the mean ratio of differentially expressed (DE) miRNAs expression and miR-103a-3p and miR-199b-5p as reference miRNAs ± standard error of the mean (SEM). Asterisks indicate a significant difference compared to the control (**** $p < 0.0001$, *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$); **(C)** Pearson correlation results of the DE miRNAs and BMI and CA125 levels. The cross stays for not significant results.

miRNA	FC	FDR
miR-1246	2.61	< 0.001
miR-150-5p	-1.81	< 0.001
miR-4454 + miR-7975	2.45	< 0.001
miR-142-3p	-1.94	< 0.001
miR-191-5p	-1.51	< 0.001
miR-15b-5p	-1.65	< 0.001
miR-126-3p	1.61	< 0.001
miR-15a-5p	-1.75	< 0.001
miR-144-3p	-2.31	0.01
miR-4516	1.68	0.01
miR-630	2.16	0.01
miR-106b-5p	-1.56	0.04

Table 2. Overview of the DE miRNAs (FC) >|1.5| in HGSOE concerning normal control. FC fold change, FDR false discovery rate.

Pathways enrichment analysis of DE miRNA target genes

To reveal the biological function of the DE miRNAs, we first identified 1568 putative DE miRNAs target genes using Ingenuity Pathway Analysis (IPA)³¹, 5779 target genes via mirDB³² and 10,652 target genes via TargetScan-Human 8.0³³. For further functional studies, we used 1024 targets that overlapped between the above databases (Fig. 2A). The protein–protein interaction (PPI) network, including 1022 nodes and 4485 edges, was generated by Cytoscape v.3.9.1³⁴. Because the network size exceeded 2000 nodes, we reduced the network to proteins that directly interact with each other (Supplementary Fig. S1). For this purpose, using a zero-order network function of the NetworkAnalyst 3.0 platform³⁵, we obtained 255 nodes and 497 edges. By ranking the PPI network nodes using eight topological analysis methods, including both local- and global-based algorithms from the cytoHubba plugin³⁶ of Cytoscape software, we found that *CCND1*, *PTEN*, *E2F1*, *STAT3*, *CDK4*, *CDK6*, *GRB2*, *RAC1*, *CREB1*, and *VEGFA* scores ranked in the top 10 (Supplementary Table S1). The highly connected hub gene *CCND1* encodes cyclin D1, which together with *CDK4* and *CDK6* kinases control G1 to S phase cell cycle progression and is frequently overexpressed in human cancers, promoting tumorigenesis^{37,38}.

Next, we used a combination of IPA, gene ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG), and WikiPathways³⁹ pathway enrichment analyses using the enrichR tool⁴⁰. For biological processes (BP), the DE miRNA target genes were enriched in the regulation of cell proliferation (FDR-corrected P value = 8.50×10^{-09}), G1/S transition of the mitotic cell cycle (FDR-corrected P value = 3.44×10^{-06}), cell cycle G1/S phase transition (FDR-corrected P value = 1.84×10^{-05}), positive regulation of cell population proliferation (FDR-corrected P value = 2.71×10^{-05}), and negative regulation of mitotic cell cycle phase transition (FDR-corrected P value = 1.49×10^{-05}) (Supplementary Fig. S2A). The IPA analysis showed the involvement of the following canonical pathways, G1/S checkpoint regulation (FDR-corrected P value = 1.37×10^{-10}), PI3K/AKT signalling (FDR-corrected P value = 2.48×10^{-10}), ovarian cancer signalling (FDR-corrected P value = 3.37×10^{-10}), and PTEN signalling (FDR-corrected P value = 6.87×10^{-10}), with molecular mechanisms of cancer (FDR-corrected P value = 2.09×10^{-11}) at the top of the list (Supplementary Figs. S2B and S3A). According to the KEGG and WikiPathways databases, the PI3K–Akt signalling pathway was the most significantly enriched pathway. Moreover, enrichment analysis showed that DE miRNAs target genes were also implicated in the EGF/EGFR signalling pathway, STAT3 signalling pathway, G1 to S cell cycle control, MAPK signalling pathway, and mTOR signalling pathway (Fig. 2B and Supplementary Fig. S3B).

Next, juxtaposed DE miRNAs target genes identified in our cohort with EOC-related from publicly available DisGeNET v7.0 database⁴¹. The 86 targets that overlapped between 1024 DE miRNA target genes and 1224 EOC-related genes were analysed in terms of known interactions using the HumanNet v3 platform⁴². Network analysis (Fig. 2C) was performed based on HumanNet-FN (functional gene network) which includes co-functional links (given by co-expression, co-essentiality, pathway database, protein domain profile associations, gene neighbourhood, and phylogenetic profile association) and protein–protein interactions. The top guide gene within this network was *E2F1* (score = 35.0), and the other guide genes within the top scores included: *CCND1*, *CDK4*, *CDK6*, *CCNE1*, *PTEN*, *ESR1*, *CCNE2*, *MYB*, and *CCND2*. *E2F1* encodes the E2F transcription factor 1, which regulates the expression of genes involved in cellular proliferation, differentiation, angiogenesis, DNA damage response, and apoptosis. Both genes, *CCND1* and *E2F1*, were upregulated in HGSOc tissue (Fig. 3A and B). Furthermore, the correlation between the expression of the selected miR-1246 and miR-150-5p (with the lowest FDR value) in serum, and that of *CCND1* and *E2F1* was analysed in tumour tissue. The results, as presented in Fig. 3C–F, demonstrate that the expressions of miR-1246 and *CCND1*, as well as miR-1246 and *E2F1*, are moderately positively correlated ($r = 0.41$ and $r = 0.58$, respectively) with a significant statistical value of ($p < 0.01$). Conversely, a moderate negative correlation was observed between miR-150-5p and *CCND1*, and miR-150-5p and *E2F1*. The correlations were also statistically significant in these instances ($p < 0.01$). Within the obtained network, we also identified downstream candidate genes that could be functionally connected to the 86 input guide genes. The top scores comprised *CDKN2A*, *RELA*, *CDK1*, *RB1*, *CDKN2A*, *EP300*, *CDKN1B*, *PRKACA* and *CREBBP* with *CDK2* and *TP53* as top candidates (score = 49.6 and score = 44.5, respectively).

Finally, we examined the miRNA profile in ovarian cancer tissues matching previously screened serum samples. We performed a small RNA-seq analysis in matched serum and tumour tissue specimens. Analysis revealed 278 differentially expressed miRNAs (DEMs) (132 upregulated and 146 downregulated) in tissue compared to 13 found in serum. As shown in Fig. 4A, only eight DE miRNAs appeared to be differentially expressed in matched tissue and serum samples. Of these, the expression of miR-1246 was significantly upregulated in serum from cancer patients, whereas it was downregulated in cancer tissue compared to the control. In addition, we found a significant over-representation of miR-1246 targets in several KEGG pathways including pathways in cancer (FDR-corrected P value = 6.37×10^{-09}), EGFR tyrosine kinase inhibitor resistance (FDR-corrected P value = 2.02×10^{-06}), PI3K–Akt signalling pathway (FDR-corrected P value = 1.06×10^{-06}), signalling pathway regulating pluripotency of stem cells (FDR-corrected P value = 6.04×10^{-05}), and MAPK signalling pathway (FDR-corrected P value = 1.76×10^{-05}) (Fig. 4B).

Construction of a Weighted Gene Co-Expression Network and correlation with clinical traits

Using the Weighted Gene Co-Expression Network (WGCNA) algorithm⁴³, we identified key miRNA modules in HGSOc. We used the normalised NanoString data to incorporate 798 miRNAs detected in 34 patients with HGSOc and 36 normal controls for the analysis (Supplementary Fig. S4A). After a series of quality assessments for the miRNAs matrix, we set the soft threshold as 5 to construct and validate the scale-free network (Supplementary Fig. S4B). Five modules were identified by setting the cut height as 0.25 to merge similar modules (Supplementary Fig. S4C). Next, the correlation between modules and clinical traits was investigated (Supplementary Fig. S5A). We identified the yellow module consisting of 26 miRNAs that were highly correlated with HGSOc, FIGO stage, tumour size, nodules and metastasis (TNM). Of the 26 miRNAs present in the yellow module, eight

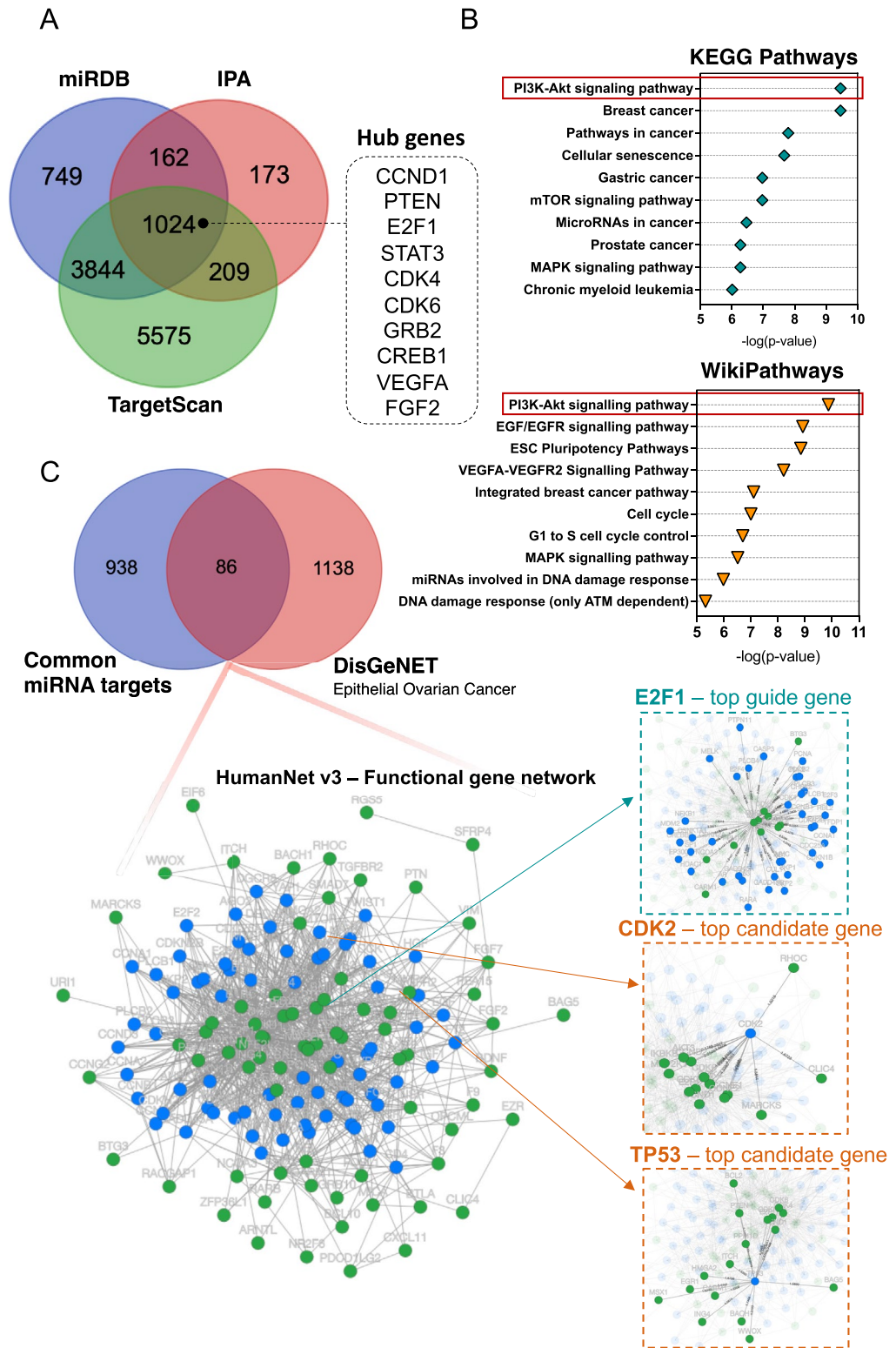


Figure 2. Functional annotation analysis of predicted miRNA-target genes. (A) Venn diagram showing the number of miRNA target genes that are shared between different databases, miRDB, TargetScan, and IPA; *miRDB* microRNA target prediction database; *IPA* ingenuity pathway analysis; (B) KEGG pathway and WikiPathways analysis of 1024 miRNA target genes. All functional annotations were performed with the hypergeometric test and Bonferroni adjustment (corrected P value (FDR) ≤ 0.05); *KEGG* Kyoto encyclopedia of genes and genomes; (C) Venn diagram showing the overlap between 1024 miRNA target genes and genes related to EOC from the DisGeNET database. Additionally, integrated functional gene network analysis of 86 miRNA target genes with the HumanNet v3 platform. Network nodes represent guide (miRNA target genes – green) genes and candidate genes (blue). Edges represent their associations. Edges guide and edges between guide genes and candidate genes are presented; *EOC* epithelial ovarian cancer; *DisGeNET* a database of gene-disease associations.

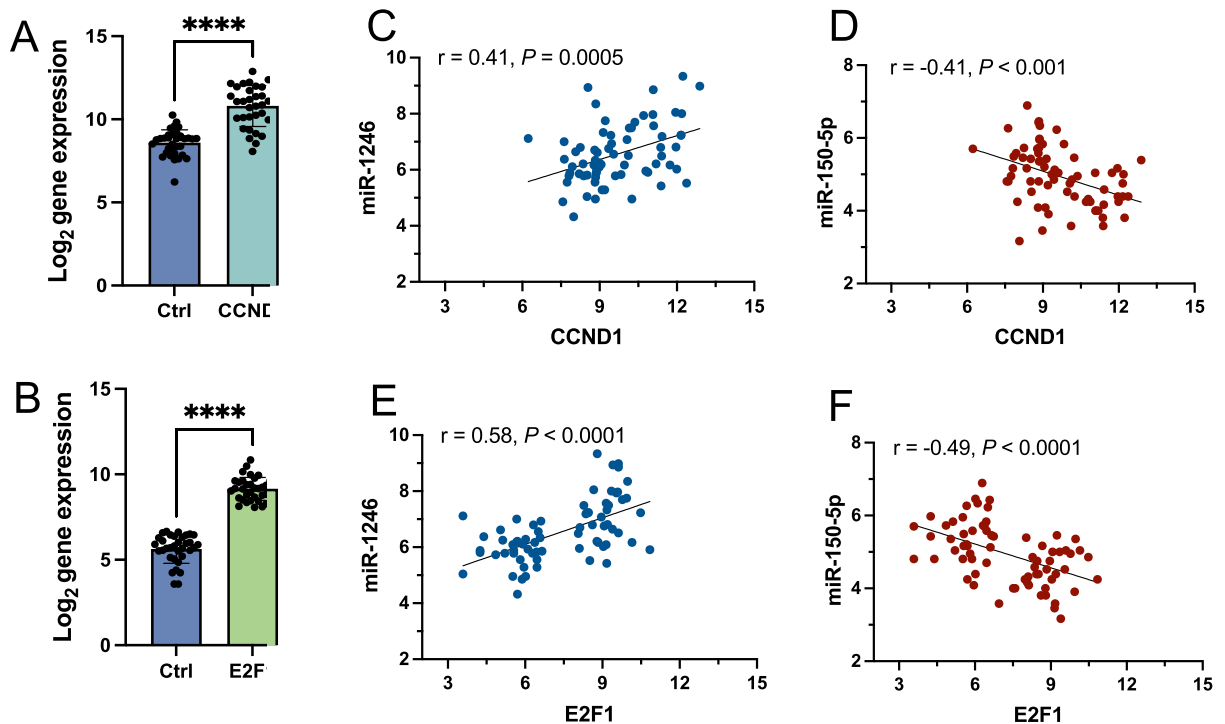


Figure 3. Expression level obtained by RNA-seq of (A) CCND1 and (B) E2F1 genes. All gene expression levels between HGSOc and non-cancer individuals were significantly different (**** $p < 0.0001$). Pearson correlation of CCND1 expression with (C) miR-1246 expression and (D) miR-150-5p expression. Pearson correlation of E2F1 expression with (E) miR-1246 and (F) miR-150-5p expression.

were previously identified as DE miRNAs in Fig. 1. Next, we analysed 188 miRNA target genes from the yellow module. The top pathway identified by KEGG and WikiPathway analyses was the PI3K-Akt signalling pathway (Supplementary Fig. S5C). Among the affected processes, we also found the EGF/EGFR signalling pathway, mTOR signalling pathway, FoxO signalling pathway, and MAPK pathway.

Evaluating circulating miRNAs as diagnostic markers for HGSOc

First, we performed receiver operating characteristic (ROC) analyses to assess the diagnostic performances of the 13 DE miRNAs to detect HGSOc. Corresponding area under the curve (AUC) values, confidence intervals (CI), sensitivity and specificity for cut-off points were calculated and are shown in Supplementary Table S2 and Supplementary Fig. S6. Four miRNAs, miR-1246, miR-150-5p, and miR-4454 + miR-7975 showed the highest AUCs reaching 0.923, 0.872, and 0.856, respectively. The occurrence, sensitivity and specificity for all DE miRNAs above 0.75 indicate that they possess good diagnostic potential.

Identification of the best combination of miRNAs for HGSOc detection

To reveal the biological function of the DE miRNAs classification power for discrimination between HGSOc and non-cancer samples, we randomly divided 70 samples into two groups, the training set (70%) and the testing set (30%) (Table 1). Based on the changed expression of miRNAs in cancer and healthy patients identified by the NanoString method, we combined two attribute selection methods, Information Gain and Correlation-based Feature Subset Selection (CfsSubsetEval), to select the strongest miRNAs candidates to develop a diagnostic classification model. Both methods were carried out in the testing set with the use of leave-one-out cross-validation (LOOVC). The Information Gain method with the Ranker Search method is based on the calculation of decreasing entropy by adding attributes and Correlation-based feature selection prioritises uncorrelated features. The top 3 attributes by InfoGainAttributeEval Algorithm, best-explaining data, are miR-1246, miR-144-3p, and miR-150-5p. The top attributes by CfsSubsetEval Algorithm, are miR-1246, miR-150-5p, miR-144-3p, miR-4454 + miR-7975, and miR-4516. The strongest attributes, which were selected as overlapping from two attribute selections, were miR-1246, miR-150-5p, and miR-144-3p. All results are listed in Supplementary Table S3.

To assess the diagnostic values of the selected above circulating miR-1246, miR-150-5p, and miR-144-3p, the multivariate logistic regression method was applied to develop the diagnostic models of the miRNAs under combination conditions. To develop stable models in the course of training, k-fold cross-validation (K = 10) was used. Based on the normalized NanoString data, two models were derived. Model 1 built on the expression of miR-1246 and miR-150-5p as independent variables, while Model 2 included miR-1246 and miR-144-3p (Supplementary Table S4). The ROC curves, and the AUCs values (Supplementary Table S5), as well as confusion matrices (Supplementary Table S6), were used to evaluate the diagnostic potential of both combination miRNAs in the training set and the testing set. The AUC was 1, the diagnostic sensitivity was 100% and the specificity

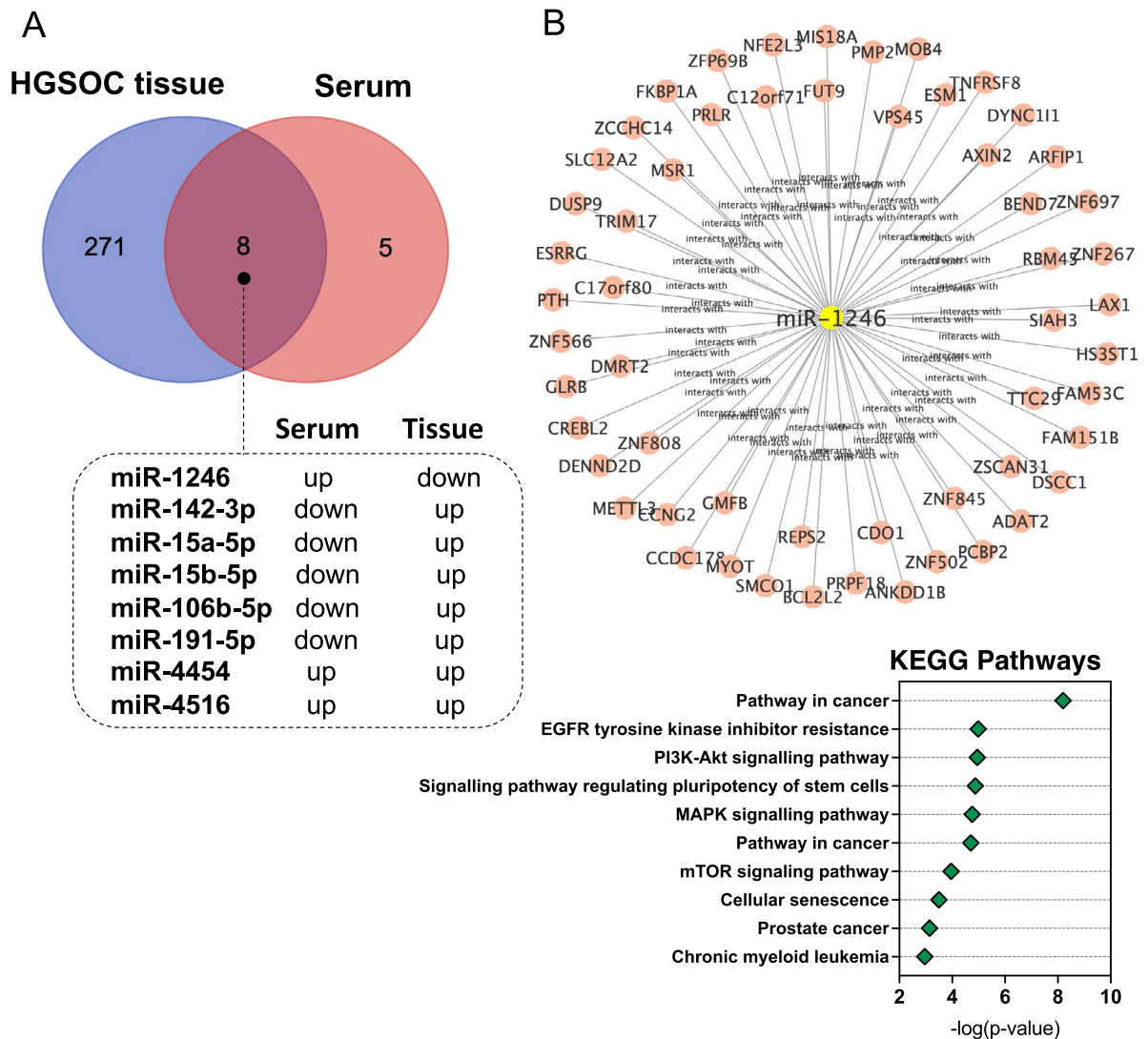


Figure 4. Comparison of miRNA expression between matched serum and tissue HGSOC samples. (A) Venn diagram of the list of DE miRNAs between serum and tissue samples. The diagram shows that eight miRNAs were significantly differentially expressed compared to healthy controls in serum and tissue samples. In addition, six out of eight miRNAs show an opposite trend of regulation, and two show a concordant trend; (B) Cytoscape network of miR-1246 and its targets and KEGG pathway analysis of miR-1246 target genes. All functional annotations were performed with the hypergeometric test and Bonferroni adjustment (corrected P value (FDR) ≤ 0.05); KEGG Kyoto encyclopedia of genes and genomes.

was 92.3% for Model 1 (Fig. 5A), and the AUC was 0.938, with a sensitivity of 92.3% and specificity of 87.5% for Model 2 in the testing set (Fig. 5B). These results indicated that Model 1 including the combination of miR-1246 and miR-150-5p showed higher sensitivity and better specificity than Model 2 and can be more effective in the diagnosis of HGSOC.

Validation of the diagnostic power of miR-1246 and miR-150-5p by RT-qPCR

Next, to validate the diagnostic values of circulating miR-1246 and miR-150-5p from Model 1, we performed RT-qPCR analysis. For this, we used an enriched group of serum samples from 42 patients with HGSOC and 46 healthy donors (Table 1). The expression levels of miR-1246 and miR-150-5p were normalized to the expression of reference miRNAs, miR-103-3p and miR-199-5p. Based on the normalized RT-qPCR data, a diagnostic classification Model 3 was developed, which was trained on a training set comprising 70% of the data and validated on a testing set (30% of the data). To develop a stable diagnostic classification model, LOOCV was used in the course of training. The results of the diagnostic performance are shown in Fig. 6 and in Supplementary Table S7, Supplementary Table S8 and Supplementary Table S9. Model 3 provided very good discriminatory power on the training set (AUC 0.997), with a sensitivity of 96.4% and specificity of 94.1% and it highly classified patients in the testing set (AUC 0.946) with a perfect sensitivity of 100% and specificity of 91.7%.

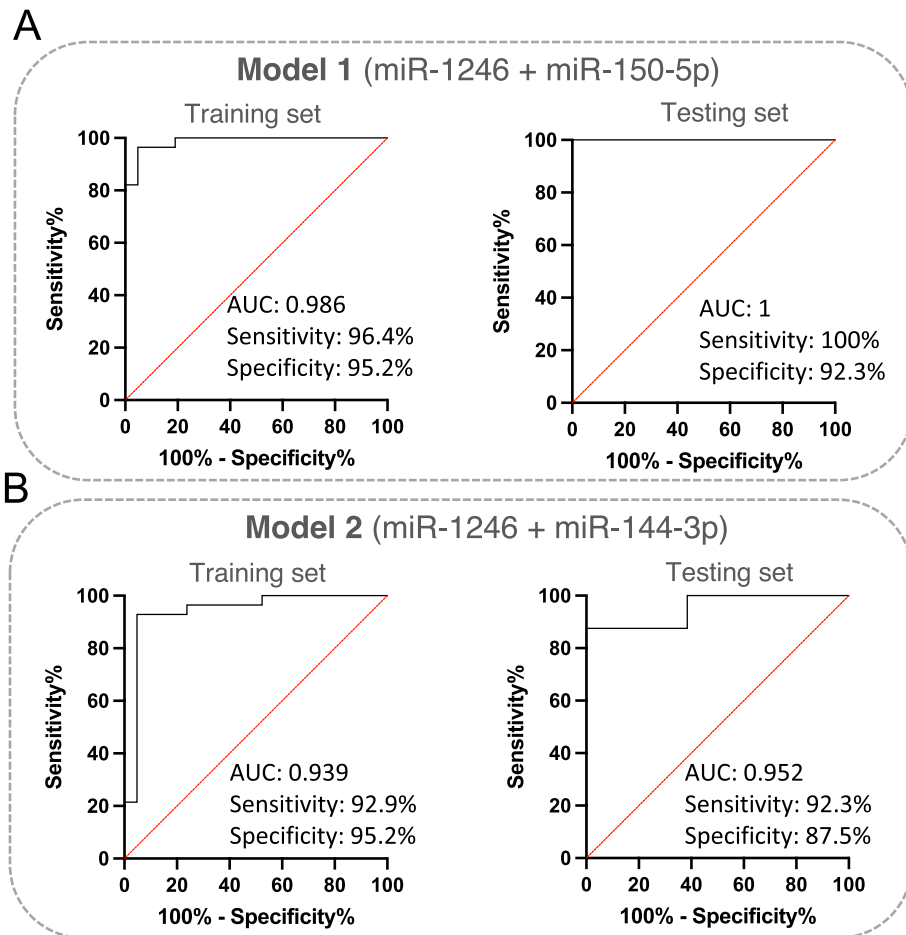


Figure 5. Development of the HGSOc diagnostic model. (A) ROC curve and AUC (area under the curve) for the training and the testing set for the diagnostic classification model based on data on the miR-1246 and miR-150-5p expression levels obtained using the NanoString platform. The graph consists of the AUC value, sensitivity and specificity corresponding to that point; (B) ROC curve and AUC (area under the curve) for the training and the testing set for the diagnostic classification model based on data on the miR-1246 and miR-144-3p expression levels obtained using the NanoString platform. The graph consists of the AUC value, sensitivity and specificity corresponding to that point.

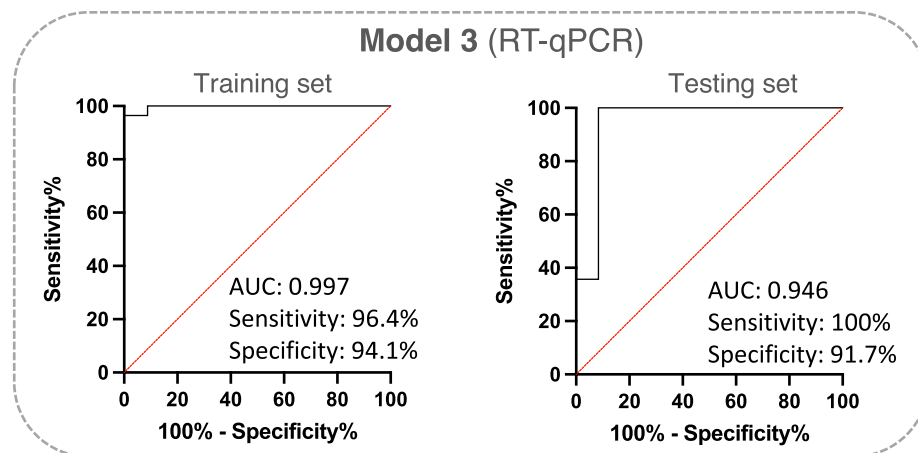


Figure 6. Validation of diagnostic power of miR-1246 and miR-150-5p by RT-qPCR method. ROC curves and AUC for the training and the testing set for the diagnostic classification model were obtained based on data obtained using the RT-qPCR technique. The graphs contain the AUC value, sensitivity and specificity corresponding to that point.

Validation of the diagnostic power of miR-1246 and miR-150-5p in an external dataset

Next, additional validation of the optimised miR-1246 and miR-150-5p signature was performed using an external dataset (GSE 106,817, 3D-Gene Human miRNA V21_1.0.0 microarrays, Toray Industries, Inc.), which contains miRNA profiles of 4046 cancer patients and 2759 healthy donors. From the available set, the data of patients with ovarian cancer ($n = 320$) and the data of matched healthy donors ($n = 320$) were selected. The whole cohort was divided into a training set (70%) and a testing set (30%). We found that Model 4, created based on miR-1246 and miR-150-5p expression data in the GEO dataset, sufficiently discriminated between OC patients and healthy donors with AUC 0.891, sensitivity 81.2%, and specificity 87.1% in the training set and AUC 0.895, sensitivity 82.3% and specificity 87.5% in the testing set (Fig. 7A, Supplementary Table S10, Table 3 and Table S11). In addition, based on the FIGO criteria, we also subdivided ovarian cancer data according to the stage, including patients with early stages of cancer (FIGO 1 and 2), which are generally difficult to detect during routine clinical screening. The results of the diagnostic performance are shown in Fig. 7B and Table 3. Based on these results, we were able to discriminate between HGSOc at all FIGO stages and healthy donors with AUC values higher than 0.88 and sensitivity in the range of 75.0–92.8% and specificity in the range of 77.5–88.7%. Given that the dataset lacks an adequate description of histopathological subtypes for each sample, however, the source publication⁴⁵

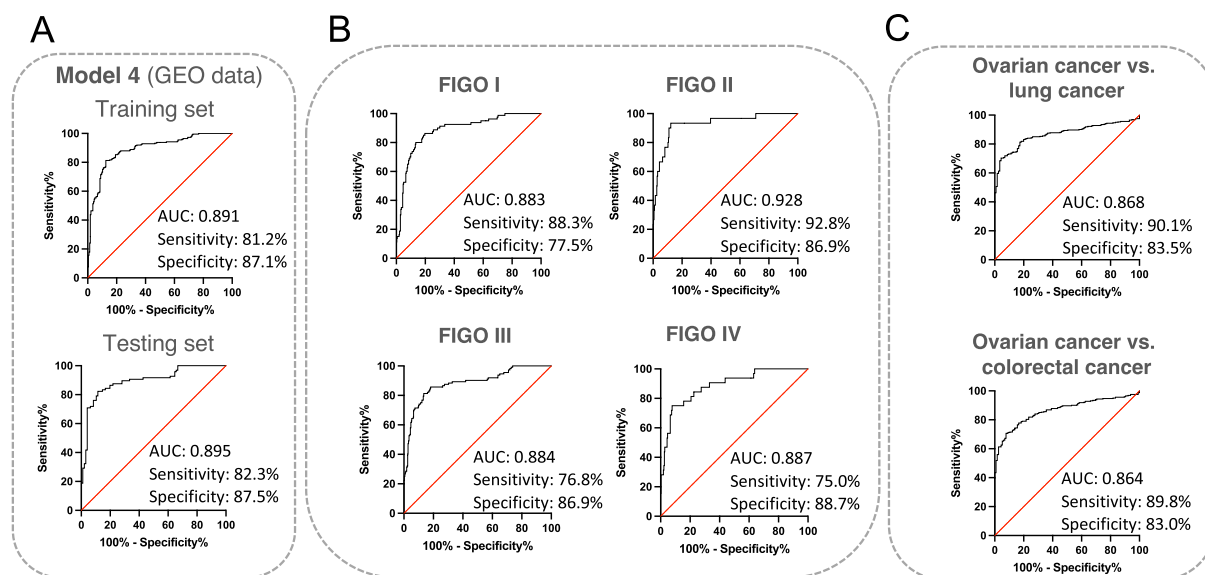


Figure 7. Validation of the HGSOC diagnostic model based on external data. (A) ROC curve and AUC for the training and the testing set for the diagnostic classification model based on public data on the miR-1246 and miR-150-5p expression levels obtained using an array technique, which is included in Gene Expression Omnibus (GEO) database. The graph consists of the AUC value, sensitivity and specificity corresponding to that point; (B) ROC curves and AUC for the diagnostic classification model depend on the ovarian cancer stage (FIGO I–IV). The graph contains the AUC value, sensitivity and specificity corresponding to that point; (C) ROC curves and AUC for the diagnostic classification model depend on the type of cancer. The graph consists of the AUC value, sensitivity and specificity corresponding to that point.

Name	miR-1246, miR-150-5p	miR-1246, miR-150-5p
Set	Training	Testing
Area under curve (AUC)	89.1%	89.5%
Confidence interval (CI) lower limit	86.1%	84.8%
CI upper limit	92.1%	94.1%
Cut-off point (Youden index)	0.56	0.56
Sensitivity	81.2%	82.3%
Specificity	87.1%	87.5%
FIGO I (n = 80), AUC (95% CI)	–	88.3%
FIGO II (n = 30), AUC (95% CI)	–	92.8%
FIGO III (n = 112), AUC (95% CI)	–	88.4%
FIGO IV (n = 32), AUC (95% CI)	–	88.7%

Table 3. Quality parameters for diagnostic classification model 4 in training and test sets for expression measurements by microarray data (GEO dataset).

includes histology distribution. Most of the samples are HGSOE (182 out of 320); the rest include clear cell, endometrioid and other epithelial and non-epithelial carcinoma.

Finally, to investigate whether the serum miR-1246 and miR-150-5p profiles can distinguish ovarian cancer from other solid cancers, we developed a model based on two miRNA expression datasets in lung and colon cancers from the GSE 106,817 dataset. As shown in Fig. 7C, the trained models were able to properly discriminate patients with ovarian cancer from patients suffering from lung cancer (AUC 0.868, sensitivity 90.1%, and specificity 83.5%) or colon cancer (AUC 0.864, sensitivity 89.8%, and specificity 83.0%).

Discussion

Despite substantial improvement in the diagnosis of heterogeneous diseases such as EOC, effective screening methods for early detection are still urgently needed. To date, several mathematical models and diagnostic algorithms have been created using histopathology and protein-based biomarker analyses in serum. However, while multivariate classification tools such as ROMA, Ova1 and Overa* have been found to be helpful for the diagnosis of EOC in general, they lack high sensitivity and specificity as well as early diagnosis potential^{27,44}. In this context circulating miRNA emerges as a particularly promising new EOC biomarker.

Here, based on miRNA expression profiles in the serum of HGSOE patients and healthy individuals, we propose a classification model that could improve the diagnosis of ovarian cancer. We found that a two-miRNA biomarker signature comprising miR-1246 and miR-150-5p can successfully differentiate between HGSOE patients and non-cancer individuals with the AUC 0.946 (95% CI 0.964–1.00). At the point of maximum classification accuracy sensitivity was 100% with a specificity of 91.7%. Our proposed diagnostic model utilizing serum miRNAs may be inexpensive, non-invasive and easily used clinically. To date, several studies have investigated using miRNAs as serum biomarkers for ovarian cancer, but there has been no combination of miR-1246 and miR-150-5p for diagnosing HGSOE. Moreover, analyses often include combined samples from various histopathological subtypes, such as serous, clear cell, endometrioid, mucinous, and other epithelial carcinomas. For example, Yokoi et al.⁴⁵ constructed a diagnostic model based on the expression levels of nine serum miRNAs (miR-1246, miR-663b, miR-4730, miR-642a-3p, miR-658, miR-486-3p, miR-1207-5p, miR-4419b, miR-6124) for discriminating ovarian cancer patients from healthy women (AUC, 0.86; S, 82%; Sp, 91%). Unfortunately, this model could not efficiently discriminate patients with benign or borderline tumours from OC patients. Kan et al.⁴⁶ identified a multivariate model combining miR-200b and miR-200c with good predictive power to discriminate patients with serous ovarian cancer, including both high- and low-grade types and healthy controls (AUC, 0.784). Similarly, multivariate analysis by Song et al.⁴⁷ revealed the diagnostic potential of miR-26b and miR-21 with an AUC of 0.916, a sensitivity of 87.5% and a specificity of 90.4%. Halvorsen et al.⁴⁸ proposed miR-200a-3p and miR-200c-3p as EOC detection biomarkers (AUC, 0.89; Sn, 84%, Sp, 83%). In another study by Todeshini et al.⁴⁹, over 230 serum samples from OC and healthy controls were screened by microarrays for miRNA marker identification. According to the results, miR-1246, miR-595, and miR-2278 were significantly over-expressed in the serum of HGSOE patients compared to healthy controls. Interestingly, ROC curve analysis revealed miR-1246 as the best predictor, with an AUC of 0.82, a sensitivity of 87%, and a specificity of 77%. In this case, the combination of three biomarkers did not increase the AUC. In our study, for the first time, we showed that a model based on combining the expression of miR-1246 and miR-150-5p indicates promising outcomes in discriminating HGSOE patients from normal controls with higher predictive values. We have also evaluated the diagnostic power of these two miRNAs in an external dataset published by Yokoi et al.⁴⁵. The evaluation model evaluation metrics (AUC, Sensitivity and Specificity) have decreased in this validation set, which might be due to the mixing of the histological OC subtypes in the dataset instead of only the HGSOE subtype, as in the discovery set. However, the resulting model is doing very well in setting the right OC diagnosis (AUC 0.895, sensitivity 82.3% and specificity 87.5%). Therefore, although the model based on miR-1246 and miR-150-5p was adjusted for HGSOE diagnosis, there is also potential in using it to recognise other subtypes of OC, however less effectively. Additionally, our data revealed that ovarian cancer patients could be sufficiently discriminated from those with lung and colorectal carcinoma. This information about analyses across various cancer types will also be helpful in the clinical application of serum miRNA panels for monitoring HGSOE.

Based on our results, we also noticed that miR-1246 was downregulated in tumour tissue samples, with a simultaneous upregulation in serum HGSOE patients. Circulating miRNA levels may reflect the condition of tumour tissue⁵⁰. Thus, it can be considered that miR-1246, as a tumour suppressor, is released from tumour cells. Increasing evidence has emerged that miRNAs are locally produced in solid tumours and secreted into several body fluids, including serum^{51,52}. Based on this evidence, it is believed that such miRNAs might be significant biomarkers for several types of cancer with high success and repeatability^{53–55}. Serum levels of miR-1246 have been found to be the most upregulated miRNA in the serum of patients with lung cancer⁵⁶. Similarly, the following studies in clinical settings have shown upregulation of miR-1246 in the serum or plasma of patients with colorectal cancer (CRC), prostate cancer, breast cancer, hepatocellular carcinoma (HCC), esophageal cancer, laryngeal squamous cell carcinoma (LSCC), and pancreatic cancer (PC)^{57–65}. The diagnostic value of miR-1246 has been validated in different neoplastic disorders. The AUC values for various cancers ranged between 0.69 and 0.97. A very promising result can be seen in hepatocellular carcinoma and breast cancer, where miR-1246 has been used as a diagnostic marker with AUC values of 0.97 and 0.967, respectively^{66,67}. The expression level of miR-150-5p in plasma samples has been shown to distinguish prostate cancer patients from healthy individuals with an AUC value of 0.817⁶⁸. Likewise, Zou et al.⁶⁹ reported the diagnostic potential of miR-150-5p in colorectal cancer with an AUC of 0.87.

In addition to miR-1246 and miR150-5p, we also identified several circulating DE miRNAs (miR-144-3p, miR-142-3p, miR-15a-5p, miR-15b-5p, miR-126-3p, miR-4454, miR-7975, miR-191-5p, miR-4516, miR-630, and miR-106b-5p), which have been previously reported to play a crucial role in HGSOE and other types of

cancers. Decreased expression of miR-150-5p, miR-106b-5p, miR-126-3p, miR-142-3p, and miR-191-5p has been observed in presurgical plasma from serous epithelial ovarian cancer⁶². Importantly, miR-150-5p together with miR-106b-5p and miR-126-3p, have shown differential expression in benign presurgical plasma. Similarly, Todeschini et al.⁴⁹ showed upregulated expression of miR-1246 along with downregulation of miR-106b-5p in miRNA profiling of serum samples from 168 HGSOC patients. In addition, they also reported significantly altered expression of miR-1246, miR-150-5p, miR-144-3p, miR-15a-5p, miR-15b-5p, miR-126-3p, and miR-106b-5p in matching HGSOC tissue. Previous research has also approved that serum levels of miR-142-3p were significantly lower in colorectal cancer (CRC) patients than in healthy controls. Furthermore, in this study, survival analysis demonstrated that CRC patients with low serum miR-142-3p levels had a lower 5-year overall survival rate⁷⁰. Similar, to HGSOC, miR-150-5p was significantly downregulated in plasma samples from prostate cancer (PC) patients compared to cancer-free controls. A similar study by Carvalho et al.⁷¹ showed a lower expression level of miR-150-5p along with miR-142-3p in exosomes from breast cancer patients, which correlated well with the stage of the tumour.

Our functional annotation analyses of predicted DE miRNA targets revealed gene enrichment in key cancer pathways relevant to tumorigenesis including, PI3K/AKT signalling pathway, EGFR signalling pathway, MAPK pathway, STAT3 signalling pathway, PTEN signalling pathway, and regulation of the cell cycle transition from G1 to S phase. EGFR signalling promotes ovarian tumorigenesis including proliferation, migration and angiogenesis, and high EGFR expression correlates with poor prognosis^{72,73}. A major downstream signalling pathway of EGFR is the phosphoinositide 3-kinase (PI3K)/protein kinase B (AKT) pathway, which plays a crucial role in OC tumorigenesis and is related to aggressive phenotypes, chemo- and radiotherapy resistance, and poor prognosis in OC^{74–76}. Furthermore, the relationship between hyperactive EGFR signalling through STAT3 and PI3K activity together promotes high-grade ovarian cancer progression to cisplatin resistance^{77,78}. By network-based analysis, we identified the top 10 hub genes, including *CCND1*, *E2F1*, *PTEN*, *STAT3*, *CDK4*, *CDK6*, and *GRB2*. Their expression might be regulated by various mechanisms, along with regulation by miRNAs, for instance, by miR-106a-5p, miR-150-5p, miR-15-5p, miR-144-3p, miR-4516 and miR-1246^{79,80}. Cyclin D1 (*CCND1*) is one of the cell cycle regulators modulating the development of malignant tumours. It has been reported that the *CCND1* gene can bind to cyclin-dependent kinase 4 (CDK4) and CDK6 to form a complex, which cooperates with the cyclin E-CDK2 complex and releases the transcription factor E2F, thus leading to an abnormal cell cycle and tumorigenesis⁸¹. PTEN is a major downstream effector of the PI3K pathway, a phosphatase capable of antagonising the PI3K/AKT pathway through the dephosphorylation of the phosphatidylinositol (3,4,5) – triphosphate (PIP3)⁸². PTEN also directly dephosphorylates AKT1, leading to the inhibition of tumorigenesis by regulating the PTEN/PIP3/PDK1-Akt signalling pathway. Leveraging the HumanNet-FN platform, we also identified top guide and candidate genes within the created networks based on overlapping putative target mRNAs identified for the serum miRNA profile and genes related to EOC by the DisGeNET database. We identified *E2F1*, which encoded E2F Transcription Factor 1, as the top guide gene. *E2F1* is well-known as a good prognostic indicator in ovarian carcinoma and its overexpression has been associated with unfavourable disease-free and overall survival^{83,84}. Recent studies have shown that the *E2F1* gene can be regulated by the PI3K/AKT pathway. AKT has been shown to directly phosphorylate E2F1, which enhances its transcriptional activity and promotes cell proliferation⁸⁵. In addition, the PI3K/AKT pathway can regulate the expression of other genes that control the cell cycle, such as cyclin D1 and p27^{Kip1}, which can, in turn, regulate E2F1 activity⁸⁶.

Our study has several limitations. First, the sample size of this study was limited. Therefore, we further performed large-scale validation to determine if the constructed model was efficacious. Second, our model was built to predict diagnosis and cannot be used to predict prognosis or recurrence. Thus, a longitudinal study may clarify the effectiveness of our model for treatment effects and timely diagnosis. In addition, because over 80% of cases were stratified into FIGO III-IV stages, the proposed model will not be sufficient for early diagnosis, however, based on the results from the external dataset, there is potential, that the model is as discriminative for high as for low FIGO stage. In summary, a large-scale study is now required to confirm the role of these miRNAs as biomarkers for HGSOC.

Given that OC is a heterogeneous disease frequently lacking symptoms at the early stage and with later metastasis stages often misinterpreted as related to digestive system malfunctions, improved diagnostic tools are needed. Tissue biopsy is the gold standard for the diagnosis of a variety of tumours. However, in the case of OC, tissue biopsies have to be avoided because puncture can cause cancer cells to disseminate into the peritoneal cavity, promoting peritoneal metastasis. We report 12 DE serum-derived miRNAs in HGSOC patients and their functional analysis. In this study, we have also developed a diagnostic model based on the expression of two serum-circulating miRNAs, which shows high sensitivity and specificity for identifying HGSOC patients. The developed diagnostic model might be an important tool in the future of gynaecologic oncology as non-invasive diagnosis support.

Methods

Study cohort

Serum samples were collected in the Clinical Hospital in Bialystok by Biobank at the Medical University of Bialystok, with high standards of strict biobanking procedures described further by Niklinski et al.⁸⁷, between 2015 and 2019. Finally, the present study involved 70 patients, including 34 patients with HGSOC and 35 non-cancer individuals for NanoString screening analysis, and 88 patients, including 42 patients with HGSOC and 46 healthy controls for RT-qPCR validation (Table 1). The serum was collected at baseline before surgery and any treatment along with the clinical information about patients' age, BMI CA125 level, FIGO stage and TNM stage. Women diagnosed with types of cancer other than high-grade serous ovarian cancer, chemotherapy or radiotherapy before serum collection, and material containing at least 50% tumour cells were excluded from

the study. According to the recommendation of the International Federation of Gynaecology and Obstetrics⁸⁸, patients were stratified into FIGO I FIGO II, FIGO III, and FIGO IV stages in the NanoString and the RT-qPCR analysis. Samples from healthy donors with no cancer history were obtained from the Clinical Research Centre, Medical University of Białystok. Serum was harvested from patients by centrifugation of blood in serological tubes and stored at -80°C until further use. Sample collection was approved by the local ethics committee of the Medical University of Białystok, Poland (approval number: R-I-002/36/2014 and APK.002.69.2020) and conducted according to the Declaration of Helsinki. All participants provided written informed consent.

Sample size estimation

Based on our previous experiments and pilot data, we calculated the minimal number of samples per experimental group (tumour or normal) to detect two-fold differences in relative expression levels between groups at true positive detection powers of 80% and 90%⁸⁹. Since tumour and normal tissue differ in terms of inter-individual variations, we used the RNASeqPower R package to apply the statistics data covering obtained real counts and coefficients of variation per group. For the NanoString nCounter miRNA data, we estimated that to obtain 80% power, we would need nine samples per group, whereas to obtain a high power of 90%, we would need 12 samples. For the smallRNA-Seq data, we estimated that to obtain 80% power, we would need 20 samples per miRNA group, whereas to obtain a high power of 90%, we would need 27 samples per miRNA group. Finally, our groups for NanoString nCounter and smallRNA-seq analyses consisted of 34 (normal tissue) and 36 (tumour tissue) samples, thus allowing for more than 90% power in any of the comparisons performed.

RNA preparation and miRNA profiling by NanoString

RNA isolation with miRNA fraction serum samples was performed using the miRNeasy Serum/Plasma Advanced Kit (Qiagen, Germany) according to the manufacturer's instructions. The analysis was performed using the nCounter[®] Analysis System (NanoString Technologies, WA, USA) and the nCounter Human v3 miRNA Panel. Briefly, as input material, three ng of isolated miRNA was used. Next, unique DNA tags were ligated onto the 3' end of each mature miRNA, followed by overnight hybridization (65°C) to nCounter Reporter and Capture probes. After hybridization, samples were placed into the nCounter Prep Station for sample purification and target/probe complex immobilization on the cartridge. For each assay, a high-density scan (555 fields of view) was performed on the nCounter Digital Analyser to count individual fluorescent barcodes and quantify target miRNA molecules in each sample.

miRNA profiling by Next Generation Sequencing

Total RNA with miRNA fraction was extracted from tissue samples using *mirVana*[™] Isolation Kit (ThermoFisher Scientific, MA, USA) according to the manufacturer's instructions. RNA concentration, purity and integrity were assessed by Qubit (Invitrogen, CA, USA) and Tape Station (Agilent Technologies, CA, USA). Small RNA-Seq libraries were constructed from 1 μg of total RNA with an RNA integrity number (RIN) > 8 using the Illumina TruSeq Small RNA Preparation Kit (Illumina, CA, USA). Indexed libraries were pooled, clustered using cBot and sequenced on the HiSeq 4000 platform, generating 50 bp single-end reads (1×50 bp). Sequencing data were processed to obtain fastq files with the bcl2fastq pipeline (Illumina, San Diego, USA), including demultiplexing and adapter trimming steps. The quality of the obtained reads was assessed using FastQC (Babraham Institute, Cambridge, United Kingdom) and multiQC⁹⁰ before the analysis, as well as after different processing steps. Reads were trimmed from adapter sequences with cutadapt⁹¹ and only reads that contained adapters were kept for further analysis. After adapter trimming reads were filtered using cutadapt for quality and length in the range of 16–28 bp. MiRNA detection was performed with miRge 2.0⁹² using bowtie for alignment and miRBase v22⁹³. Raw miRNA counts were used as input for DE miRNA analysis. MiRNA count data were filtered for lowly expressed miRNAs of less than five raw counts in the smallest library, taking into account corresponding CPM values (Counts per Million). Counts were normalized using a weighted trimmed mean of the log expression ratios (trimmed mean of M values – TMM)⁹⁴. Data were further transformed for linear modelling using voom⁹⁵. Linear modelling and empirical Bayes moderation were applied using the limma package to assess the differential expression of miRNAs⁹⁶. P values were corrected for multiple comparisons with FDR. The results plots were created in the R environment and basic statistical functions.

qPCR

miRNA extraction from serum samples was performed as described previously. Reference miRNAs with stable expression across all samples, hsa-miR-103-3p and hsa-miR-199b-5p, were selected based on the NanoString data using the NormFinder algorithm⁹⁷. The 13 DE miRNAs and two reference miRNAs were profiled using the miRCURY LNA SYBR Green PCR Kit (Qiagen, Germany). The miRCURY LNA RT Kit (Qiagen, Germany) was used for the reverse transcription reaction. The reaction was conducted with the primers as shown in Table S12. The temperature profile of the qPCR reaction was as follows: 2 min at 95°C and 50 cycles: 10 s at 95°C and 60 s at 56°C . Amplification was performed using the LightCycler 480 (Roche, Switzerland). Subsequently, PCR threshold cycles (C_t) of the tested miRNAs and reference miRNAs were determined for the tested samples and the calibrator. The relative expression for each miRNA was calculated with PCR efficiency correction⁹⁸. Efficiency (E) was calculated from the slopes of the calibration curve according to the equation: $E = 10^{-1/\text{slope}}$. Reactions with amplification efficiency below 1.6 were removed. The relative expression ratio of a target miRNA was computed based on its PCR efficiencies (E) and the C_t value difference (Δ) of unknown group samples (test) versus the control group (ΔC_t control-test). The relative calculation was based on the MEAN C_t of the experimental group.

miRNA target gene prediction and pathway analysis

The identification of DE miRNA target genes was performed using Ingenuity Pathway Analysis Software³¹ (IPA, Qiagen Inc. <https://qiagenbioinformatics.com/products/ingenuity-pathway> analysis), mirDB database³² (<https://mirdb.org>), and TargetScanHuman 8.0 database³³ (https://www.targetscan.org/vert_80/). Gene Ontology Biological Process and KEGG Pathway analyses were conducted with the ClusterProfile R package⁹⁹. IPA was used to perform the core analysis to identify canonical pathways. Over-representation analyses were performed using hypergeometric tests under $\alpha=0.05$ (P -values corrected with FDR). To construct a PPI network, we used STRING (<https://string-db.org/>; v.11.5). Genes with a confidence score ≥ 0.4 were chosen to build a network model visualized by Cytoscape v.3.9.1. Nine topological algorithms in plug-in *cytoHubba*³⁶, consisting of “MCC”, “MNC”, “Degree”, “Bottle Neck”, “EcCentricity”, “Closeness”, “Stress”, and “Radiality” were selected to identify the hub genes in PPI analysis. Analyses of functional interaction networks were based on the HumanNet v3 platform by applying the HumenNet-FN (functional gene network) mode (<https://www.inetbio.org/humannet/>)⁴².

WGCNA

Normalized counts according to the nSolver protocol were used as the starting point of the analysis. Counts were transformed to adjust for heteroscedasticity with log2 and precision weights based on the mean–variance trend. Hierarchical clustering was performed based on the transformed data (Supplementary Fig. S4A). The discovery and analysis of miRNA co-expression modules in the patient’s serum based on NanoString nCounter data were performed with Weighted Gene Coexpression Network Analysis (WGCNA)¹⁰⁰ by application of the WGCNA R libraries¹⁰¹. Outliers were removed based on standardised connectivity. Pearson correlation was used in the signed network construction with soft threshold selection ($R^2 > 0.8$) (Supplementary Fig. S4B, C). Modules were merged based on a remarkably high eigengene similarity correlation threshold (0.95). Gene modules to trait relationships were evaluated with Spearman correlation.

Diagnostic model development

The normalised NanoString counts and RT-qPCR data were used for model development. First, both datasets were split into a discovery set (70%) and a validation set (30%) (Table 1). The splits have been random and have been done, as all further calculations in R version 3.6.1^{102,103}. The attributes were selected using Waikato Environment for Knowledge Analysis (WEKA) version 3.8.3. (c) 1999–2018 The University of Waikato, Hamilton, New Zealand. Two independent algorithms InfoGainAttributeEval and Correlation-based Feature Subset Selection (CfsSubsetEval), were used on the training set to select the best classifiers. Feature selection via information gain using InfoGainAttributeEval was based on the calculation of decreasing entropy by adding attributes and selection attributes that most strongly reduce entropy. The CfsSubsetEval method evaluated the worth of a subset of attributes by considering the individual predictive ability of each feature along with the degree of redundancy between them. Both methods were carried out on the testing set using the LOOCV^{104,105}. A multivariate logistic regression model was built using a training set. Training of the model was performed using repeated ($n = 3$) k-fold-cross-validation ($k = 10$) in R version 3.6.1 (R Core Team (2013) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna. <https://www.R-project.org/>)¹⁰³. To validate the model, the ROC curve and the AUC were calculated by the pROC package¹⁰⁶. The cut-off point for the ROC curve was calculated by the Youden index. The confusion matrix including information about TP (true positives), TN (true negatives), FP (false positives), and FN (false negatives) has been prepared. The evaluation of model classification has been based on the testing dataset. The choice of miRNAs was evaluated using the independent, external and publicly available dataset GSE1089 from the Gene Expression Omnibus (GEO) database (data from: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE106817>), which contains miRNA profiles of 4089 cancer patients and 2759 healthy donors. These external data were generated using the 3D-Gene Human miRNA V21_1.0.0 (Toray Industries, Inc.) microarray platform. Patients with ovarian cancer ($n = 320$) were filtered from the dataset. Furthermore, 320 patients were randomly selected from the control dataset to balance the groups. Most of the population used for the validation of our model consisted of high-grade serous ovarian cancer (HGSOC) ($n = 182$). Other samples include clear cell ($n = 64$), endometrioid ($n = 43$), other epithelial carcinoma ($n = 17$), and non-epithelial carcinoma ($n = 13$). As the datasets include data from another platform, the new model has been trained according to the above workflow. The evaluation also followed the same steps as previously described. In addition, according to the FIGO criteria, the GEO dataset was subdivided based on stage, including patients with FIGO I ($n = 80$), FIGO II ($n = 30$), FIGO III ($n = 112$), and FIGO IV ($n = 32$). A logistic regression model was developed for data for each stage according to the above workflow. The GEO dataset also includes patients with other types of cancers. Based on these data, the model was developed and examined based on the AUC calculated for datasets combined from ovarian cancer patients and lung cancer patients or colorectal cancer.

Statistical analyses

Statistical analyses were performed using GraphPad Prism 9 (v.9.3.1) software. The Wilcoxon rank sum test (equivalent to the Mann–Whitney U test) was used to investigate the differences in BMI and age between the HGSOC patients’ group and the healthy donors. Raw miRNA data were analysed using nSolver Software v. 4.0 (NanoString Technologies, WA, USA). For technical variations, code-set content normalisation was performed relative to the ligation controls. Ratios were calculated by specifying the healthy volunteers’ samples as a baseline. Unpaired Mann–Whitney tests were performed to identify differences between patients and healthy controls. Correction for multiple testing was performed with False Discovery Rate (FDR) according to Benjamin-Hochberg. The Pearson correlation coefficient (r) was used to estimate the correlation between the identified DE miRNAs and clinical parameters.

Ethics approval and consent to participate

All procedures of human samples were conducted after the approval of the local ethics committee of the Medical University of Białystok, Poland (approval number: R-I-002/36/2014 and APK.002.69.2020) and according to the Declaration of Helsinki. All patients provided written informed consent for their participation in the study and their identities have been anonymised.

Data availability

Datasets generated by the NanoString platform are deposited on GEO: GSE235525. Other datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Received: 13 August 2023; Accepted: 18 October 2023

Published online: 07 November 2023

References

1. Siegel, R. L., Miller, K. D., Fuchs, H. E. & Jemal, A. Cancer statistics, 2022. *CA Cancer J. Clin.* **72**, 7–33 (2022).
2. Nakamura, K. *et al.* Clinical relevance of circulating cell-free microRNAs in ovarian cancer. *Mol. Cancer* **15**, 48 (2016).
3. Penyige, A. *et al.* Circulating miRNA profiling in plasma samples of ovarian cancer patients. *Int. J. Mol. Sci.* **20**, 4533 (2019).
4. Sölétormos, G. *et al.* Clinical use of cancer biomarkers in epithelial ovarian cancer: Updated guidelines from the European group on tumor markers. *Int. J. Gynecol. Cancer* **26**, 43–51 (2016).
5. Rt, F. *et al.* Ovarian cancer early detection by circulating CA125 in the context of anti-CA125 autoantibody levels: Results from the EPIC cohort. *Int. J. Cancer* **142**, 1355–1360 (2018).
6. Mehri, J. S. *et al.* Diagnostic value of tumor biomarkers CA125 and CA72–4 in differentiation of epithelial ovarian cancer and endometrioma. *Biomed. Res.* **29**, 697–701 (2018).
7. Rg, M. *et al.* Evaluation of the diagnostic accuracy of the risk of ovarian malignancy algorithm in women with a pelvic mass. *Obstet. Gynecol.* **118**, 280–288 (2011).
8. Nolen, B. M. & Lokshin, A. E. Protein biomarkers of ovarian cancer: The forest and the trees. *Future Oncol.* **8**, 55–71 (2012).
9. Whitwell, H. J., Blyuss, O., Menon, U., Timms, J. F. & Zaikin, A. Parenclitic networks for predicting ovarian cancer. *Oncotarget* **9**, 22717–22726 (2018).
10. Whitwell, H. J. *et al.* Improved early detection of ovarian cancer using longitudinal multimarker models. *Br. J. Cancer* **122**, 847–856 (2020).
11. Moore, R. G. *et al.* A novel multiple marker bioassay utilizing HE4 and CA125 for the prediction of ovarian cancer in patients with a pelvic mass. *Gynecol. Oncol.* **112**, 40–46 (2009).
12. Montagnana, M. *et al.* The ROMA (risk of ovarian malignancy algorithm) for estimating the risk of epithelial ovarian cancer in women presenting with pelvic mass: Is it really useful?. *Clin. Chem. Lab. Med.* **49**, 521–525 (2011).
13. Rein, B. J. D. *et al.* Potential markers for detection and monitoring of ovarian cancer. *J. Oncol.* **2011**, 1–17 (2011).
14. Schwarzenbach, H., Nishida, N., Calin, G. A. & Pantel, K. Clinical relevance of circulating cell-free microRNAs in cancer. *Nat. Rev. Clin. Oncol.* **11**, 145–156 (2014).
15. Huang, W. MicroRNAs: Biomarkers, diagnostics, and therapeutics. *Methods Mol. Biol.* **1617**, 57–67 (2017).
16. Friedman, R. C., Farh, K. K. H., Burge, C. B. & Bartel, D. P. Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res.* **19**, 92–105 (2009).
17. Vishnoi, A. & Rani, S. MiRNA biogenesis and regulation of diseases: An overview. In *Methods in Molecular Biology* Vol. 1509 (ed. Rani, S.) 1–10 (Humana Press Inc., 2017).
18. Vienberg, S., Geiger, J., Madsen, S. & Dalgaard, L. T. MicroRNAs in metabolism. *Acta Physiol.* **219**, 346–361 (2017).
19. Tiwari, A., Mukherjee, B. & Dixit, M. MicroRNA key to angiogenesis regulation: MiRNA biology and therapy. *Curr. Cancer Drug Targets* **18**, 266–277 (2017).
20. Bautista-Sánchez, D. *et al.* The promising role of miR-21 as a cancer biomarker and its importance in RNA-based therapeutics. *Mol. Ther. Nucleic Acids* **20**, 409–420 (2020).
21. Xiang, Y., Tian, Q., Guan, L. & Niu, S. S. The dual role of miR-186 in cancers: Oncomir battling with tumor suppressor miRNA. *Front. Oncol.* **10**, 33 (2020).
22. Ryu, J. K. *et al.* Aberrant microRNA-155 expression is an early event in the multistep progression of pancreatic adenocarcinoma. *Pancreatol.* **10**, 66–73 (2010).
23. Qin, W., Ren, Q., Liu, T., Huang, Y. & Wang, J. MicroRNA-155 is a novel suppressor of ovarian cancer-initiating cells that targets CLDN1. *FEBS Lett.* **587**, 1434–1439 (2013).
24. Iorio, M. V. *et al.* MicroRNA signatures in human ovarian cancer. *Cancer Res.* **67**, 8699–8707 (2007).
25. Kim, S. *et al.* Serum exosomal miRNA-145 and miRNA-200c as promising biomarkers for preoperative diagnosis of ovarian carcinomas. *J. Cancer* **10**, 1958–1967 (2019).
26. Wang, W. *et al.* Five serum microRNAs for detection and predicting of ovarian cancer. *Eur. J. Obstet. Gynecol. Reprod. Biol.* **X3**, 100017 (2019).
27. Kumar, V., Gupta, S., Varma, K. & Sachan, M. MicroRNA as biomarker in ovarian cancer management: Advantages and challenges. *DNA Cell Biol.* **39**, 103–124 (2020).
28. Mirahmadi, Y. *et al.* MicroRNAs as biomarkers for early diagnosis, prognosis, and therapeutic targeting of ovarian cancer. *J. Oncol.* **2021**, 1–25 (2021).
29. Staicu, C. E. *et al.* Role of microRNAs as clinical cancer biomarkers for ovarian cancer: A short overview. *Cells* **9**, 169 (2020).
30. Schmid, G. *et al.* Expression and promotor hypermethylation of miR-34a in the various histological subtypes of ovarian cancer. *BMC Cancer* **16**, 102 (2016).
31. Krämer, A., Green, J., Pollard, J. & Tugendreich, S. Causal analysis approaches in ingenuity pathway analysis. *Bioinformatics* **30**, 523–530 (2014).
32. Chen, Y. & Wang, X. miRDB: An online database for prediction of functional microRNA targets. *Nucleic Acids Res.* **48**, D127–D131 (2020).
33. Agarwal, V., Bell, G. W., Nam, J.-W. & Bartel, D. P. Predicting effective microRNA target sites in mammalian mRNAs. *eLife* **4**, e05005 (2015).
34. Shannon, P. *et al.* Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).
35. Zhou, G. *et al.* NetworkAnalyst 3.0: A visual analytics platform for comprehensive gene expression profiling and meta-analysis. *Nucleic Acids Res.* **47**, W234–W241 (2019).
36. Chin, C.-H. *et al.* cytoHubba: Identifying hub objects and sub-networks from complex interactome. *BMC Syst. Biol.* **8**(Suppl 4), S11 (2014).

37. Dall'Acqua, A. *et al.* Inhibition of CDK4/6 as therapeutic approach for ovarian cancer patients: Current evidences and future perspectives. *Cancers (Basel)* **13**, 3035 (2021).
38. Qie, S. & Diehl, J. A. Cyclin D1, cancer progression and opportunities in cancer treatment. *J. Mol. Med. (Berl.)* **94**, 1313–1326 (2016).
39. Martens, M. *et al.* WikiPathways: Connecting communities. *Nucleic Acids Res.* **49**, D613–D621 (2021).
40. Kuleshov, M. V. *et al.* Enrichr: A comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* **44**, W90–97 (2016).
41. Piñero, J. *et al.* The DisGeNET knowledge platform for disease genomics: 2019 update. *Nucleic Acids Res.* **48**, D845–D855 (2020).
42. Kim, C. Y. *et al.* HumanNet v3: An improved database of human gene networks for disease research. *Nucleic Acids Res.* **50**, D632–D639 (2022).
43. Langfelder, P. & Horvath, S. WGCNA: An R package for weighted correlation network analysis. *BMC Bioinform.* **9**, 559 (2008).
44. Abdalla, N. *et al.* Does the risk of ovarian malignancy algorithm provide better diagnostic performance than HE4 and CA125 in the presurgical differentiation of adnexal tumors in polish women?. *Dis. Mark.* **2018**, 5289804 (2018).
45. Yokoi, A. *et al.* Integrated extracellular microRNA profiling for ovarian cancer screening. *Nat. Commun.* **9**, 4319 (2018).
46. Kan, C. W. *et al.* Elevated levels of circulating microRNA-200 family members correlate with serous epithelial ovarian cancer. *BMC Cancer* **12**, 1–9 (2012).
47. Song, K.-W., Zhang, Q.-G., Tan, W.-B. & Fang, Y.-N. Diagnostic significance of serum miR-26b and miR-21 expressions in ovarian cancer and their associations with clinicopathological characteristics and prognosis of patients. *Eur. Rev. Med. Pharmacol. Sci.* **24**, 1697–1703 (2020).
48. Halvorsen, A. R. *et al.* Evaluation of prognostic and predictive significance of circulating MicroRNAs in ovarian cancer patients. *Dis. Mark.* **2017**, 1–9 (2017).
49. Todeschini, P. *et al.* Circulating miRNA landscape identifies miR-1246 as promising diagnostic biomarker in high-grade serous ovarian carcinoma: A validation across two independent cohorts. *Cancer Lett.* **388**, 320–327 (2017).
50. Wang, J., Zhang, K.-Y., Liu, S.-M. & Sen, S. Tumor-associated circulating MicroRNAs as biomarkers of cancer. *Molecules* **19**, 1912–1938 (2014).
51. Volinia, S. *et al.* A microRNA expression signature of human solid tumors defines cancer gene targets. *Proc. Natl. Acad. Sci.* **103**, 2257–2261 (2006).
52. Calin, G. A. & Croce, C. M. MicroRNA signatures in human cancers. *Nat. Rev. Cancer* **6**, 857–866 (2006).
53. Zheng, H. *et al.* Plasma miRNAs as diagnostic and prognostic biomarkers for ovarian cancer. *PLoS One* **8**, e77853 (2013).
54. Schneider, A. *et al.* Tissue and serum microRNA profile of oral squamous cell carcinoma patients. *Sci. Rep.* **8**, 675 (2018).
55. Chhatrya, B. *et al.* Comparison of tumour and serum specific microRNA changes dissecting their role in pancreatic ductal adenocarcinoma: A meta-analysis. *BMC Cancer* **19**, 1175 (2019).
56. Yang, F. *et al.* MiR-1246 promotes metastasis and invasion of A549 cells by targeting GSK-3 β -mediated Wnt/ β -catenin pathway. *Cancer Res. Treat.* **51**, 1420–1429 (2019).
57. Guo, S. *et al.* Exosomes derived from Fusobacterium nucleatum-infected colorectal cancer cells facilitate tumour metastasis by selectively carrying miR-1246/92b-3p/27a-3p and CXCL16. *Gut* **10**, (2020).
58. Salah, M. *et al.* Detection of miR-1246, miR-23a and miR-451 in sera of colorectal carcinoma patients: A case-control study in Cairo University hospital. *Afr. Health Sci.* **20**, 1283–1291 (2020).
59. Bhagirath, D. *et al.* microRNA-1246 is an exosomal biomarker for aggressive prostate cancer. *Cancer Res.* **78**, 1833–1844 (2018).
60. Zhang, Z. *et al.* Exosomal miR-1246 and miR-155 as predictive and prognostic biomarkers for trastuzumab-based therapy resistance in HER2-positive breast cancer. *Cancer Chemother. Pharmacol.* **86**, 761–772 (2020).
61. Ahmed, E. K., Fahmy, S. A., Effat, H. & Abdel-Wahab, A. H. Circulating miR-210 and miR-1246 as potential biomarkers for differentiating hepatocellular carcinoma from metastatic tumors in the liver. *J. Med. Biochem.* **38**, 109–117 (2019).
62. Chen, S. *et al.* Expression and diagnostic value of miR-497 and miR-1246 in hepatocellular carcinoma. *Front. Genet.* **12**, 66306 (2021).
63. Hoshino, I. *et al.* Usefulness of serum miR-1246/miR-106b ratio in patients with esophageal squamous cell carcinoma. *Oncol. Lett.* **20**, 1–1 (2020).
64. Huang, Q. *et al.* Lack of miR-1246 in small extracellular vesicle blunts tumorigenesis of laryngeal carcinoma cells by regulating Cyclin G2. *IUBMB Life* **72**, 1491–1503 (2020).
65. Xu, Y.-F., Hannafon, B. N., Zhao, Y. D., Postier, R. G. & Ding, W.-Q. Plasma exosome miR-196a and miR-1246 are potential indicators of localized pancreatic cancer. *Oncotarget* **8**, 77028–77040 (2017).
66. Moshiri, F. *et al.* Circulating miR-106b-3p, miR-101-3p and miR-1246 as diagnostic biomarkers of hepatocellular carcinoma. *Oncotarget* **9**, 15350–15364 (2018).
67. Cui, X. *et al.* Breast cancer identification via modeling of peripherally circulating miRNAs. *PeerJ* **6**, e4551 (2018).
68. Paunescu, I. A. *et al.* Biomarker potential of plasma MicroRNA-150-5p in prostate cancer. *Medicina (Kaunas)* **55**, 564 (2019).
69. Zou, S.-L. *et al.* Downregulation of serum exosomal miR-150-5p is associated with poor prognosis in patients with colorectal cancer. *Cancer Biomark.* **26**, 69–77 (2019).
70. Gao, W., Pang, D. & Yu, S. Serum level of miR-142-3p predicts prognostic outcome for colorectal cancer following curative resection. *J. Int. Med. Res.* **47**, 2116–2125 (2019).
71. Carvalho, T. M. *et al.* MicroRNAs miR-142-5p, miR-150-5p, miR-320a-3p, and miR-4433b-5p in serum and tissue: Potential biomarkers in sporadic breast cancer. *Front. Genet.* **13**, 865472 (2022).
72. Siwak, D. R. *et al.* Targeting the epidermal growth factor receptor in epithelial ovarian cancer: Current knowledge and future challenges. *J. Oncol.* **2010**, e568938 (2009).
73. Mak, V. C. *et al.* p85 β alters response to EGFR inhibitor in ovarian cancer through p38 MAPK-mediated regulation of DNA repair. *Neoplasia* **23**, 718–730 (2021).
74. Lheureux, S., Gourley, C., Vergote, I. & Oza, A. M. Epithelial ovarian cancer. *Lancet* **393**, 1240–1253 (2019).
75. Ediriweera, M. K., Tennekoon, K. H. & Samarakoon, S. R. Role of the PI3K/AKT/mTOR signaling pathway in ovarian cancer: Biological and therapeutic significance. *Semin. Cancer Biol.* **59**, 147–160 (2019).
76. Mabuchi, S., Kuroda, H., Takahashi, R. & Sasano, T. The PI3K/AKT/mTOR pathway as a therapeutic target in ovarian cancer. *Gynecol. Oncol.* **137**, 173–179 (2015).
77. Yue, P. *et al.* Hyperactive EGF receptor, Jaks and Stat3 signaling promote enhanced colony-forming ability, motility and migration of cisplatin-resistant ovarian cancer cells. *Oncogene* **31**, 2309–2322 (2012).
78. Colomiere, M. *et al.* Cross talk of signals between EGFR and IL-6R through JAK2/STAT3 mediate epithelial-mesenchymal transition in ovarian carcinomas. *Br. J. Cancer* **100**, 134–144 (2009).
79. Chen, M., Zang, M. & Guo, X. MicroRNA-106a-5p functions as an oncogene via regulating PTEN in breast cancer cells. *Int. J. Clin. Exp. Med.* **12**, 8044–8056 (2019).
80. Sagar, S. K. miR-106b as an emerging therapeutic target in cancer. *Genes Dis.* **9**, 889–899 (2022).
81. Quan, L.-L. *et al.* Expression of cyclin D1 gene in ovarian cancer and effect of silencing its expression on ovarian cancer cells based on the oncomine database. *Bioengineered* **12**, 9290–9300 (2021).
82. Bu, L. *et al.* PTEN suppresses tumorigenesis by directly dephosphorylating Akt. *Signal Transduct. Target Ther.* **6**, 1–3 (2021).

83. Reimer, D. *et al.* Clinical relevance of E2F family members in ovarian cancer—an evaluation in a training set of 77 patients. *Clin. Cancer Res.* **13**, 144–151 (2007).
84. Farra, R., Dapas, B., Grassi, M., Benedetti, F. & Grassi, G. E2F1 as a molecular drug target in ovarian cancer. *Expert Opin. Ther. Targets* **23**, 161–164 (2019).
85. Al-Bahlani, S. M. *et al.* Correlation of expression of Akt1 and E2F1 and their phosphorylated forms in breast cancer patients with clinicopathological parameters. *J. Mol. Histol.* **52**, 621–633 (2021).
86. Arcaro, A. & Guerreiro, A. S. The phosphoinositide 3-kinase pathway in human cancer: Genetic alterations and therapeutic implications. *Curr. Genom.* **8**, 271–306 (2007).
87. Niklinski, J. *et al.* Systematic biobanking, novel imaging techniques, and advanced molecular analysis for precise tumor diagnosis and therapy: The polish MOBIT project. *Adv. Med. Sci.* **62**, 405–413 (2017).
88. Berek, J. S., Kehoe, S. T., Kumar, L. & Friedlander, M. Cancer of the ovary, fallopian tube, and peritoneum. *Int. J. Gynaecol. Obstet.* **143**(Suppl 2), 59–78 (2018).
89. Hart, S. N., Therneau, T. M., Zhang, Y., Poland, G. A. & Kocher, J.-P. Calculating sample size estimates for RNA sequencing data. *J. Comput. Biol.* **20**, 970 (2013).
90. Ewels, P., Magnusson, M., Lundin, S. & Källér, M. MultiQC: Summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **32**, 3047–3048 (2016).
91. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**, 10–12 (2011).
92. Lu, Y., Baras, A. S. & Halushka, M. K. miRge 2.0 for comprehensive analysis of microRNA sequencing data. *BMC Bioinform.* **19**, 275 (2018).
93. Griffiths-Jones, S., Saini, H. K., van Dongen, S. & Enright, A. J. miRBase: Tools for microRNA genomics. *Nucleic Acids Res.* **36**, D154–158 (2008).
94. Robinson, M. D. & Oshlack, A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* **11**, R25 (2010).
95. Law, C. W., Chen, Y., Shi, W. & Smyth, G. K. voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol.* **15**, R29 (2014).
96. Law, C. W. *et al.* RNA-seq analysis is easy as 1–2–3 with limma, glimma and edge R. *F1000Research* **5**, 1408 (2018).
97. Andersen, C. L., Jensen, J. L. & Ørntoft, T. F. Normalization of real-time quantitative reverse transcription-PCR data: A model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. *Cancer Res.* **64**, 5245–5250 (2004).
98. Pfaffl, M. W. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res.* **29**, e45 (2001).
99. Yu, G., Wang, L.-G., Han, Y. & He, Q.-Y. clusterProfiler: An R package for comparing biological themes among gene clusters. *OMICS* **16**, 284–287 (2012).
100. Langfelder, P. & Horvath, S. WGCNA: An R package for weighted correlation network analysis. *BMC Bioinform.* **9**, 559 (2008).
101. Russo, P. S. T. *et al.* CEMiTool: A bioconductor package for performing comprehensive modular co-expression analyses. *BMC Bioinform.* **19**, 56 (2018).
102. R Core Team R: A language and environment for statistical computing. R foundation for statistical computing, Vienna. <https://www.R-project.org> (2021).
103. Kuhn, M. Building predictive models in R using the caret package. *J. Stat. Softw.* **28**, 1–26 (2008).
104. Rafalo, M. Cross validation methods: Analysis based on diagnostics of thyroid cancer metastasis. *ICT Express* **8**, 183–188 (2022).
105. Wainer, J. & Cawley, G. Nested cross-validation when selecting classifiers is overzealous for most practical applications. *Expert Syst. Appl.* **182**, 115222 (2021).
106. Robin, X. *et al.* pROC: An open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinform.* **12**, 77 (2011).

Author contributions

Conception and design: M.N., A.E., A.K.; Acquisition of data: M.N., A.Z., A.B., M.K., J.S., J.R., P.K.; Analysis and interpretation of data: M.N., A.E., K.C., A.S.; Drafting of the manuscript: M.N., A.E.; Critical revision of the manuscript for important intellectual content: A.S., J.S., P.K., M.M., A.K.; Statistical analysis: A.E., K.C., M.N.; Supervision: A.S., M.M., A.S.; The author(s) read and approved the final manuscript.

Funding

This study was supported by funds from the Ministry of Education and Science of Poland, within the project “Excellence Initiative—Research University”, and the Ministry of Health of Poland, within the project “Center for Artificial Medicine at the Medical University of Bialystok”. “This research was conducted within the project which has received funding from the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 754432 and the Polish Ministry of Education and Science, from the financial resources for science in 2018–2023 granted for the implementation of an international co-financed project”.



Competing interests

Patent number: 17P46787PL00, applicant: Medical University of Bialystok, inventors: Magdalena Niemira (MN), Anna Erol (AE), Adam Kretowski (AK), status – active, title: Panel of miRNA biomarkers for diagnosis of ovarian cancer, method for in vitro diagnosis of ovarian cancer, uses of panel of miRNA biomarkers for in vitro diagnosis of ovarian cancer and test for in vitro diagnosis of ovarian cancer. The other authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-023-45317-7>.

Correspondence and requests for materials should be addressed to M.N.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023