A gene expression signature of epithelial ovarian cancer in the blood.

Shahan Mamoor, MS¹

¹shahanmamoor@gmail.com

East Islip, NY USA

Epithelial ovarian cancer is the most lethal gynecologic cancer in the developed world¹. A major factor contributing to the fatality rate associated with epithelial ovarian cancer is late diagnosis² resulting from the lack of suitable diagnostic and screening measures, like the mammogram in breast cancer³ and the prostate specific antigen (PSA) blood test for human prostate cancer^{4,5}. Radiologic screening can be cost-prohibitive⁶. We mined published microarray data⁷ to discover genes expressed in the blood that could distinguish patients with epithelial ovarian cancer from those without. We identified a core set of 31 genes, 10 of which independently adequately distinguished control patients from patients with epithelial ovarian cancer. 7 out of 10 of these genes were also differentially expressed in tumor tissue from epithelial ovarian cancer^{8,9}. These data can serve as a basis for the development of a blood-based, quantitative PCR diagnostic screening tool for the detection of epithelial ovarian cancer at the population level.

Keywords: epithelial ovarian cancer, targeted diagnostics in epithelial ovarian cancer, standards for the diagnosis and treatment of cancer in women.

The 5-year relative survival rate for patients with epithelial ovarian cancer diagnosed at stage I is 89%, but 17% at stage IV¹⁰. Early diagnosis is hampered by the failure of the 21st century medical establishment to develop and implement adequate diagnostics for the prevention of late-stage ovarian cancer. We mined published microarray data to identify genes whose expression was most significantly different between the blood of patients with epithelial ovarian cancer as compared to non-affected controls, identifying a set of 10 genes whose expression distinguished the blood of patients with epithelial ovarian cancer from others. We integrated findings from the peripheral blood with comparative transcriptome analysis of tumor tissues from patients with epithelial ovarian cancer, finding that across two datasets, out of 10 genes that served as a fingerprint of epithelial ovarian cancer in the blood, we observed differential expression of 7 out of 10 genes in tumors of the ovary. This set of genes in total can be developed as a diagnostic panel for screening of epithelial ovarian cancer at the population level, with radiologic assessment for patients who are deemed at risk based on gene expression profile of their peripheral blood.

Methods

We used datasets GSE31682⁷, GSE146553⁸ and GSE124766⁹ for this differential gene expression analysis of the blood and tumor tissues in human epithelial ovarian cancer. GSE31682 was generated using ABI Human Genome Survey Microarray Version 2 technology with blood from $n=20$ control patients and $n=48$ patients with epithelial ovarian cancer; analysis was performed using platform GPL2986. GSE146553⁸ was generated using Affymetrix Human Gene 1.0 ST Array with $n=3$ for normal ovarian tissue and $n=40$ for tumors from patients with high-grade serous ovarian cancer; analysis performed using GPL6244. GSE124766⁹ was generated using Agilent-014850 Whole Human Genome Microarray 4x44K G4112F with $n=3$ of for normal ovarian tissue and $n=8$ for tumors from patients with high-grade serous ovarian cancer; analysis performed using platform GPL6480. The Benjamini and Hochberg method of p -value adjustment was used for ranking of differential expression but raw p -values were used to assess statistical significance of global or multiplexed differential expression. Log-transformation of data was auto-detected, and the NCBI generated category of platform annotation was used. A statistical test was performed to evaluate whether the expression of genes in this panel was significantly between the blood of control patients and the blood of patients with epithelial ovarian cancer using a two-tailed, unpaired t-test with Welch's correction. We used PRISM for all statistical analyses of differential gene expression in human breast cancer (Version 8.4.0)(455).

Results

We mined published microarray data to discover genes that served as peripheral blood markers of epithelial ovarian cancer, the most lethal gynecologic malignancy, through mining published microarray data to compare global gene expression between control patients and those with epithelial ovarian cancer. We integrated these findings with comparative transcriptome analysis of tumor tissues from patients with epithelial ovarian cancer as compared to the normal ovary.

Identification of an epithelial ovarian cancer gene expression signature in the peripheral blood.

We identified 31 genes in the peripheral blood whose expression was most significantly different between control patients and patients with epithelial ovarian cancer (Figure 1 and Table 1), 10 of which adequately distinguished patients with EOC from others with minimal overlap in mRNA expression levels (Figure 2). These ten genes included beta-1,3-galactosyltransferase 2 (B3GALT2: $p=9.51E-45$), lengsin, lens protein with glutamine synthetase domain (LGSN: $p=1.44E-31$), B-cell CLL/lymphoma 9-like (BCL9L: $p=5.8E-28$), regulator of G-protein signaling 16 (RGS16: $p=3.97E-27$), chromosome 8 open

reading frame 37 (C8ORF37: $p=9.12\text{E-}26$), BTB domain containing 3 (BTBD3: $p=1.34\text{E-}23$), zinc finger protein 546 (ZNF546: $p=-1.41\text{E+}01$), E74 like ETS transcription factor 1 (ELF1: $p=1.77\text{E-}22$), caudal type homeobox 4 (CDX4: $p=1.13\text{E-}16$), and polypyrimidine tract binding protein 3 (PTBP3: $p=1.18\text{E-}16$) (Figure 3 and Table 1). The expression of each of these genes was significantly different from that of non-affected controls (Figure 3; $p<0.0001$ for all comparisons).

A core set of 7 genes is also differentially expressed in the tumor tissues of patients with epithelial ovarian cancer.

We used published microarray datasets to evaluate whether the set of genes we identified as differentially expressed in the blood was conserved across tissue or were specifically biomarkers in the peripheral blood. In one dataset, we identified differential expression of BCL9L ($p=1.99\text{E-}02$), RGS16 ($p=3.7\text{E-}03$), ELF1 ($p=2.37\text{E-}02$) and PTBP3 ($p=1.38\text{E-}02$), while in the other we identified differential expression of B3GALT2 ($p=1.37\text{E-}09$), BCL9L ($p=2.85\text{E-}02$), RGS16 ($p=9.94\text{E-}03$), C8ORF37 ($p=2.47\text{E-}04$), and ZNF546 ($p=8.15\text{E-}06$) (Tables 2 and 3, respectively) when comparing ovarian tumor tissues to the normal ovary. In total, we identified 7 genes - B3GALT2, BCL9L, RGS16, ELF1, PTBP3, C8ORF37, and ZNF546 that were differentially expressed, both in the blood and in the primary tumors of patients with epithelial ovarian cancer.

Thus, we identified a set of 10 genes whose expression in the blood, at the mRNA level, resembled a fingerprint of the most lethal gynecological cancer: B3GALT2, LGSN, BCL9L, RGS16, C8ORF37, BTBD3, ZNF546, ELF1, CDX4, and PTBP3. 7 of these genes, B3GALT2, BCL9L, RGS16, ELF1, PTBP3, C8ORF37, and ZNF546, were also among the genes most differentially expressed in the ovarian tumors of patients with epithelial ovarian cancer, demonstrating conserved transcriptional modulation of a core set of genes in the blood and tumor tissues of patients with epithelial ovarian cancers, arrived at in an unbiased fashion and through analyses at the transcriptome level.

Discussion

Relative survival for patients with epithelial ovarian cancer diagnosed at stage I is 89%, but 17% at stage IV¹⁰. There is no gold standard for the screening and prevention of epithelial ovarian cancer in the United States or elsewhere, likely due to monetary concerns regarding a simple radiologic screening based on magnetic resonance imaging (MRI)^{11,12}; the sensitivity and specificity of CA125, a blood test used to detect ovarian cancer has been described as poor¹³.

We provide here a set of 10 genes whose expression, in the peripheral blood, served to adequately distinguish the blood of patients with epithelial ovarian cancer from that of non-affected control patients. Oncotype DX is a panel of 21 genes utilized in the evaluation of human breast cancer to provide a metric known as a recurrence score; Oncotype DX is an reverse transcriptome polymerase chain reaction (RT-PCR) based assay that is utilized in molecular pathology¹⁴. We propose a subset of the 31 genes that can be similarly utilized, either in a probe-based quantitative real-time PCR (qRT-PCR) format, in mini-chips pre-printed with complementary probes to hybridize and detect messenger RNAs for each gene from the peripheral blood, or to be counted absolutely alongside a set of controls by standardized machinery, similar to nCounter technology^{15,16}. Of these options, the first, a qRT-PCR assay for the screening of epithelial ovarian cancer in the general population, could be developed and implemented with most ease.

We provided here a resource describing a molecular fingerprint of epithelial ovarian cancer in the blood and a subset of these genes to be utilized for the early detection of the most lethal gynecologic cancer in the developed world.

References

1. Berkenblit, A. and Cannistra, S.A., 2005. Advances in the management of epithelial ovarian cancer. *The Journal of reproductive medicine*, 50(6), p.426.
2. Torre, L.A., Trabert, B., DeSantis, C.E., Miller, K.D., Samimi, G., Runowicz, C.D., Gaudet, M.M., Jemal, A. and Siegel, R.L., 2018. Ovarian cancer statistics, 2018. *CA: a cancer journal for clinicians*, 68(4), pp.284-296.
3. Gøtzsche, P.C. and Jørgensen, K.J., 2013. Screening for breast cancer with mammography. *Cochrane database of systematic reviews*, (6).
4. Smith, R.A., Cokkinides, V. and Eyre, H.J., 2006. American Cancer Society guidelines for the early detection of cancer, 2006. *CA: a cancer journal for clinicians*, 56(1), pp.11-25.
5. Carroll, P., Coley, C., McLeod, D., Schellhammer, P., Sweat, G., Wasson, J., Zietman, A. and Thompson, I., 2001. Prostate-specific antigen best practice policy—part I: early detection and diagnosis of prostate cancer. *Urology*, 57(2), pp.217-224.
6. Black, W.C., Gareen, I.F., Soneji, S.S., Sicks, J.D., Keeler, E.B., Aberle, D.R., Naeim, A., Church, T.R., Silvestri, G.A., Gorelick, J. and Gatsonis, C., 2014. Cost-effectiveness of CT screening in the National Lung Screening Trial. *N Engl J Med*, 371, pp.1793-1802.
7. Pils, D., Tong, D., Hager, G., Obermayr, E., Aust, S., Heinze, G., Kohl, M., Schuster, E., Wolf, A., Sehouli, J. and Braicu, I., 2013. 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*, 13(1), p.178.
8. Zhang, W., Klinkebiel, D., Barger, C.J., Pandey, S., Guda, C., Miller, A., Akers, S.N., Odunsi, K. and Karpf, A.R., 2020. Global DNA hypomethylation in epithelial ovarian cancer: passive demethylation and association with genomic instability. *Cancers*, 12(3), p.764.
9. Hoffmann, K., Berger, H., Kulbe, H., Thillainadarasan, S., Mollenkopf, H.J., Zemojtel, T., Taube, E., Darb-Esfahani, S., Mangler, M., Sehouli, J. and Chekerov, R., 2020. Stable expansion of high-grade serous ovarian cancer organoids requires a low-Wnt environment. *The EMBO journal*, 39(6), p.e104013.

10. Baldwin, L.A., Huang, B., Miller, R.W., Tucker, T., Goodrich, S.T., Podzielinski, I., DeSimone, C.P., Ueland, F.R., van Nagell, J.R. and Seamon, L.G., 2012. Ten-year relative survival for epithelial ovarian cancer. *Obstetrics & Gynecology*, 120(3), pp.612-618.
11. Kubik-Huch, R.A., Dörffler, W., Von Schulthess, G.K., Marincek, B., Köchli, O.R., Seifert, B., Haller, U. and Steinert, H.C., 2000. Value of (18F)-FDG positron emission tomography, computed tomography, and magnetic resonance imaging in diagnosing primary and recurrent ovarian carcinoma. *European radiology*, 10(5), pp.761-767.
12. deSouza, N.M., O'Neill, R., McIndoe, G.A., Dina, R. and Soutter, W.P., 2005. Borderline tumors of the ovary: CT and MRI features and tumor markers in differentiation from stage I disease. *American Journal of Roentgenology*, 184(3), pp.999-1003.
13. Moss, E.L., Hollingworth, J. and Reynolds, T.M., 2005. The role of CA125 in clinical practice. *Journal of clinical pathology*, 58(3), pp.308-312.
14. Schmittgen, T.D. and Livak, K.J., 2008. Analyzing real-time PCR data by the comparative C T method. *Nature protocols*, 3(6), p.1101.
15. Kulkarni, M.M., 2011. Digital multiplexed gene expression analysis using the NanoString nCounter system. *Current protocols in molecular biology*, 94(1), pp.25B-10.
16. Malkov, V.A., Serikawa, K.A., Balantac, N., Watters, J., Geiss, G., Mashadi-Hosseini, A. and Fare, T., 2009. Multiplexed measurements of gene signatures in different analytes using the Nanostring nCounter™ Assay System. *BMC research notes*, 2(1), p.80.

Figure 1: A gene expression signature of epithelial ovarian cancer in the peripheral blood.

A heatmap graphically representing a molecular fingerprint of epithelial ovarian cancer (EOC) in the peripheral blood relative to that of non-affected controls.

Figure 2: A set of genes representing the most differentially expressed when comparing the blood of patients with epithelial ovarian cancer with that of non-affected controls.

Messenger RNA values for each of the 31 genes we identified as among the most differentially expressed in the blood of patients with epithelial ovarian cancer (red) is depicted here, superimposed with that of non-affected controls (green). Underlined is the set of 10 genes whose expression in the blood most adequately distinguished that of patients with epithelial ovarian cancer from non-affected controls.

Figure 3: 10 genes, each of which whose expression adequately distinguishes the blood of patients with epithelial ovarian cancer from the blood of patients with the blood of non-affected controls with minimal overlap.

The mRNA expression level in controls (each graph, left) and in patients with epithelial ovarian cancers (each graph, right “EOC”) is graphically represented with mean mRNA expression level marked and the result of a statistical test evaluating the significance of difference in gene expression between controls and patients with epithelial ovarian cancer, a p-value, listed above each graph.

Rank	ID	p-value	t	B	Gene	Gene name
1	209321	9.51E-45	-3.22E+01	86.25313344	B3GALT2	beta-1,3-galactosyltransferase 2
3	101231	1.44E-31	-2.02E+01	59.87528152	LGSN	lengsin, lens protein with glutamine synthetase domain
6	164408	5.8E-28	1.76E+01	52.20651739	BCL9L	B-cell CLL/lymphoma 9-like
7	173025	3.97E-27	-1.71E+01	50.41108069	RGS16	regulator of G-protein signaling 16
8	716436	9.12E-26	1.62E+01	47.47024404	C8orf37	chromosome 8 open reading frame 37
10	119684	1.34E-23	-1.48E+01	42.75837083	BTBD3	BTB domain containing 3
11	212552	1.67E-22	-1.41E+01	40.35832025	ZNF546	zinc finger protein 546
12	115907	1.77E-22	1.41E+01	40.30453588	ELF1	E74 like ETS transcription factor 1
19	179099	1.13E-16	1.08E+01	27.48160281	CDX4	caudal type homeobox 4
20	173247	1.18E-16	-1.07E+01	27.4386333	PTBP3	polypyrimidine tract binding protein 3
21	152891	1.28E-16	-1.07E+01	27.35851452	GCNT1	glucosaminyl (N-acetyl) transferase 1, core 2
22	105160	1.69E-16	-1.07E+01	27.09230808	CFAP65	cilia and flagella associated protein 65
23	165464	1.99E-16	-1.06E+01	26.93336827	IL36A	interleukin 36, alpha
32	136788	2.18E-14	-9.51	22.39079577	ALX4	ALX homeobox 4
33	149817	2.21E-14	-9.51	22.37775826	SLC1A7	solute carrier family 1 member 7
34	153981	2.22E-14	9.51	22.37259623	KIAA0319L	KIAA0319 like
35	184009	2.36E-14	-9.49	22.31174753	PRX	periaxin
36	178622	2.45E-14	-9.48	22.27531714	TNS4	tensin 4
38	116958	3.48E-14	-9.40	21.93530597	SLC5A4	solute carrier family 5 member 4
41	168815	7.23E-14	9.23	21.2271928	USP20	ubiquitin specific peptidase 20
45	200995	1.55E-13	9.05	20.48593826	SEPT2	septin 2
46	110153	1.61E-13	-9.05	20.4541893	SLC35E1	solute carrier family 35 member E1
47	208235	1.62E-13	-9.04	20.44577497	PRSS3	protease, serine 3
48	157342	1.75E-13	9.03	20.37301433	RSRP1	arginine and serine rich protein 1
49	144634	1.83E-13	9.02	20.33001635	GTPBP2	GTP binding protein 2
58	184050	3.98E-13	-8.84	19.57409982	PMPCA	peptidase, mitochondrial processing alpha subunit
62	139154	8.64E-13	8.66	18.82291589	IQCK	IQ motif containing K
65	112379	9.97E-13	-8.62	18.68404522	C19orf54	chromosome 19 open reading frame 54
66	208216	1.03E-12	-8.61	18.65032371	STAP2	signal transducing adaptor family member 2
67	119175	1.06E-12	8.61	18.62388635	GLYR1	glyoxylate reductase 1 homolog
206	176640	7.2E-11	-7.63	14.53392015	EVPL	envoplakin

Table 1: A gene expression signature of epithelial ovarian cancer in the human blood.

The rank of differential expression transcriptome-wide, p-value with respect to global differential expression, t, a moderated t-statistic, B, the log-odds of differential expression between the groups compared (blood of patients with EOC and blood of non-affected controls), the gene and gene name are listed here.

Rank	ID	p-value	t	B	Gene	Gene name	% DE
5692	A_24_P134816	1.99E-02	2.6926633	-3.6053	BCL9L	B-cell CLL/lymphoma 9-like	86.1%
2735	A_23_P320578	3.7E-03	3.6084022	-1.9703	RGS16	regulator of G-protein signaling 16	93.3%
6177	A_24_P78590	2.37E-02	2.596026	-3.77417	ELF1	E74 like ETS transcription factor 1	85%
4857	A_24_P295452	1.38E-02	2.8907543	-3.25531	PTBP3	polypyrimidine tract binding protein 3	88.2%

Table 2: Genes differentially expressed in the human blood of patients with epithelial ovarian cancer are also differentially expressed in the tumor tissues of patients with epithelial ovarian cancer.

The rank of differential expression transcriptome-wide, p-value with respect to global differential expression, t, a moderated t-statistic, B, the log-odds of differential expression between the groups compared, the gene, gene name and percent differential expression (% DE) are listed here. Data here are from GSE124766.

Rank	ID	p-value	t	B	Gene	Gene name	% DE
69	7923034	1.37E-09	-7.6165181	11.8289566	B3GALT2	beta-1,3-galactosyltransferase 2	99.8%
6234	7952116	2.85E-02	2.2642738	-4.1087258	BCL9L	B-cell CLL/lymphoma 9-like	78.6%
4309	7922717	9.94E-03	2.6935347	-3.1761288	RGS16	regulator of G-protein signaling 16	85.2%
1344	8151898	2.47E-04	-3.9866155	0.2428702	C8orf37	chromosome 8 open reading frame 37	95.4%
509	8028770	8.15E-06	-5.0452627	3.4837592	ZNF546	zinc finger protein 546	98.3%

Table 3: Genes differentially expressed in the human blood of patients with epithelial ovarian cancer are also differentially expressed in the tumor tissues of patients with epithelial ovarian cancer.

The rank of differential expression transcriptome-wide, p-value with respect to global differential expression, t, a moderated t-statistic, B, the log-odds of differential expression between the groups compared, the gene, gene name and percent differential expression (% DE) are listed here. Data here are from GSE146553.