Published in final edited form as:

Gynecol Oncol. 2007 December ; 107(3): 526–531.

COMBINING MULTIPLE SERUM TUMOR MARKERS IMPROVES DETECTION OF STAGE I EPITHELIAL OVARIAN CANCER

Zhen Zhang^{*,1}, Yinhua Yu², Feng-Xi Xu², Andrew Berchuck³, Carolien van Haaften-Day⁴, Laura J. Havrilesky³, Henk W. A. de Bruijn⁴, Ate G. J. van der Zee⁵, Robert P. Woolas⁶, Ian J. Jacobs⁷, Steven Skates⁸, Daniel W. Chan¹, and Robert C. Bast Jr²

¹Center for Biomarker Discovery, Department of Pathology, Johns Hopkins Medical Institutions, CRB-II 3M04, 1550 Orleans Street, Baltimore, MD, 21231 ²Anderson Cancer Center, 1515 Holcombe Blvd, Houston, TX 77030 ³Duke University Medical Center, Durham, NC 27710 ⁴Royal Hospital for Women, Sydney, Australia ⁵University Hospital Groningen, Hanzeplein 1, 9700 RB Groningen, the Netherlands ⁶Saint Mary's Hospital, Portsmouth, UK ⁷Bart's and The London, Queen Mary's School of Medicine, London, UK ⁸Massachusetts General Hospital and Harvard Medical School, Boston, MA

Abstract

Objective—Currently available tumor markers for ovarian cancer are still inadequate in both sensitivity and specificity to be used for population-based screening. Artificial neural network (ANN) as a modeling tool has demonstrated its ability to assimilate information from multiple sources and to detect subtle and complex patterns. In this paper, an ANN model was evaluated for its performance in detecting early stage epithelial ovarian cancer using multiple serum markers.

Methods—Serum specimens collected at four institutions in the US, the Netherlands, and the United Kingdom were analyzed for CA125II, CA72-4, CA15-3, and macrophage colony stimulating factor (M-CSF). The four tumor marker values were then used as inputs to an ANN derived using a training set from 100 apparently healthy women, 45 women with benign conditions arising from the ovary, and 55 invasive epithelial ovarian cancer patients (including 27 stage I/II cases). A separate validation set from 27 apparently healthy women, 56 women with benign conditions, and 35 women with various types of malignant pelvic masses was used to monitor the ANN's performance during training. An independent test dataset from 98 apparently healthy women and 52 early stage epithelial ovarian cancer patients (38 stage I and 4 stage II invasive cases and 10 stage I borderline ovarian tumor cases) was used to evaluate the ANN.

Results—ROC analysis confirmed the overall superiority of the ANN-derived composite index over CA125II alone ($p=0.0333$). At a fixed specificity of 98%, the sensitivities for ANN and CA125II alone were 71% (37/52) and 46% (24/52) ($p=0.047$), respectively, for detecting early stage epithelial ovarian cancer, and 71% (30/42) and 43% (18/42) ($p=0.040$), respectively, for detecting invasive early stage epithelial ovarian cancer.

Conclusions—The combined use of multiple tumor markers through an ANN improves the overall accuracy to discern healthy women from patients with early stage ovarian cancer. Analysis of multiple markers with an ANN may be a better choice than the use of CA125II alone in a two-step

* Corresponding author. Center for Biomarker Discovery, Johns Hopkins Medical Institutions, CRB-II 3M04, 1550 Orleans Street, Baltimore, MD 21231, Email: zzhang7@jhmi.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

approach for population screening in which a secondary test such as ultrasound is used to keep the overall specificity at an acceptable level.

Keywords

Ovarian cancer; tumor marker; early detection; multivariate model; artificial neural network

INTRODUCTION

Despite decades of clinical research with multi-modality therapy, the overall prognosis of ovarian cancer is still largely determined by whether a patient is diagnosed at an early stage when the cancer has not spread beyond the ovary (1). At present there is no effective screening strategy for ovarian cancer that will detect disease at an early stage. Screening modalities evaluated to date have included transvaginal sonography (TVS) and the use of the serum tumor marker CA 125. Three major studies have evaluated TVS for early detection and have yielded a positive predictive value of less than 10%, at the margin of the specificity required for effective screening (2-4). CA 125, a high molecular weight mucin (MUC 16) (5) was first detected with a radioimmunoassay in patients with advanced ovarian cancer (6). CA 125 levels can be elevated 10-21 months prior to conventional diagnosis (7-9) and are increased in 58% of patients with stage I ovarian cancer (10-12). The specificity of individual values of CA 125 is not adequate for effective screening, particularly in a premenopausal population where endometriosis, adenomyosis and retrograde menstruation can increase CA 125 levels. Using a two-stage strategy, however, a combination of CA 125 followed by sonography does attain specificity in excess of 99.6% (13). Using this strategy, Jacobs, et al, randomized 21,962 postmenopausal women more than 45 years of age in the United Kingdom to a control group or to a screened group. CA 125 was measured annually for three years. An elevated CA125 level >30 units/mL prompted TAU and surgery was undertaken if the TAU demonstrated a pelvic mass. Among 10,985 women screened, 29 operations were performed to detect 6 cancers, providing a positive predictive value of 21. Median survival in the screened group (72.9 months) was significantly greater ($p=0.0112$) than that in the control group (41.8 months).

Use of CA 125 as an initial stage in a 2-step screening strategy is limited by the fact that CA 125 is expressed by only 80% of epithelial ovarian cancers. The sensitivity of such an approach is further reduced due to the inability of sonography to detect a mass that has yet to become appreciable. Multiple serum markers might be utilized to improve sensitivity. Over the years a large number of markers for ovarian cancer have been reported to complement CA 125 (14). Sensitivity has been improved by 5 to 15%, but specificity has inevitably been reduced. In a study of 89 sera from patients with stage I ovarian cancer, use of 3 markers in combination (CA 125, OVX1 and M-CSF) detected 84% of cancers, whereas CA 125 alone detected 69%. Specificity, however, declined from 99% to 84% with the combination (11,12).

Novel mathematical analyses might improve sensitivity using multiple markers without sacrificing specificity. A number of different models have been suggested to combine multiple tumor markers. Statistical models such as multivariate logistic regression and nonlinear classification methods such as classification and regression tree (CART) have also been tested in the literature (15). Specificity of CA 125 could be further improved by following the trend of CA 125 with an algorithm that estimated the risk of ovarian cancer (16,17).

Artificial neural network (ANN) analysis as a statistical modeling tool has demonstrated the ability to assimilate information from multiple sources and detect subtle and complex patterns. There have been a large number of reports on the use of ANN for clinical diagnosis (18), staging and prognosis (19), and management of cancers (20). In a previous study, an ANN-based composite diagnostic index derived using a panel of four serum markers, CA 125II, CA 72-4,

CA 15-3, and lipid-associated sialic acid (LASA), was evaluated for its ability to discriminate malignant from benign pelvic masses (21). Using data from 255 patients with benign or malignant pelvic masses, the ANN index increased test specificity by approximately 20% over that of using CA125II alone, while maintaining a comparable level of sensitivity. Among an additional group of 196 healthy donors, the index had a specificity of 100%. A total of 18 patients with stage I or II epithelial ovarian cancer were included in the test set of this study. For these patients, CA 125II at the cutoff value of 35 U/ml had a sensitivity of 78% and a specificity of 68% for all benign cases. The ANN index, on the other hand, at a cutoff that yielded the same specificity as CA125II for all benign cases, provided a sensitivity of 88.9% for the stage I and II epithelial ovarian cancers.

Due to the small number of early stage ovarian cancer cases (n=18) in the previous study (21), statistical significance was not attained in comparing the performance of the ANN index against that of CA125 alone. In this paper, we studied sera from a larger number of early stage epithelial ovarian cancer patients to extend our evaluation of artificial neural network analysis for combining multiple serum tumor markers to form a single-valued index for detecting early stage ovarian cancer with both improved sensitivity and specificity compared to CA125II alone.

SUBJECTS AND METHODS

Study Population

A total of 468 serum samples were obtained from the Duke University Medical Center, Durham, N.C., St. Bartholomew's Hospital, London, United Kingdom and the Gronigen University Hospital, Gronigen, The Netherlands. The collection included: 115 specimens from patients with epithelial ovarian cancer (72 stage I, 7 stage II, and 36 late stage), 225 specimens from healthy donors, 101 specimens from women with benign conditions arising from the ovary, and 27 specimens from women with other types of malignant pelvic masses. Of these a "test set" described below included 52 early stage epithelial ovarian cancer patients with 38 stage I and 14 stage II invasive cases and 10 stage I borderline tumors. Of these 14 were serous, 14 were mucinous, 10 were endometrioid and 4 were clear cell for the invasive cases and 2 were serous and 8 were mucinous for the borderline tumors.

Sample Preparation and Tumor Marker Measurements

Blood had been permitted to clot and serum was promptly separated. All samples were stored at -70°C and thawed immediately prior to assay. The CA 125II, CA 15-3 and CA 72-3 assays were performed using radioimmunoassay kits (Centocor, Malvern PA) according to the manufacturer's instructions. M-CSF was assayed by methods previously described (22).

ANN Model Derivation

The 468 samples were allocated prospectively into three separate data sets. The training data set consisted of sera from healthy donors, women with benign conditions arising from ovary, and early and late stage epithelial ovarian cancer patients. Marker values for these sera were directly involved in the modification of connection weights in the ANN model during training. To gauge the progress during neural network training and to avoid model over-fitting, a second data set was used as a supplement for in-training validation. Finally, to test the ability of the derived ANN composite index to detect early stage epithelial ovarian cancer in a population, the independent test data set used only results from healthy donors and early stage cancer patients.

The marker values are preprocessed numerically before being used as inputs to the ANN. It involves logarithmic transformation and a "physiologically-based" normalization step. In this

step, marker values are truncated to within a range most relevant to the differentiation of patients before normalization (e.g., two patients' CA125 at 500 U/ml and 1000 U/ml would both be considered very high and be truncated and treated as the same for ANN prediction purpose).

The common practice to use data from multiple sites to derive statistical predictive models is to pool all samples together and divide them through random selection into a training set and a test set. This allows the model being derived to capture all possible data variations from different sites. By doing so, however, the test data and training data are artificially made to be guaranteed to have the same distribution. As a result, the model performance estimated from such test data tends to appear better than what the model will do when it is being placed in actual "field use." In this study, in order to have a conservative assessment of the performance of an ANN-derived algorithm, the early stage epithelial ovarian cancer specimens from a single independent site were set aside as part of the independent test samples and were not involved in any model development processes.

The Multilayer Perceptron (MLP) architecture of neural network was used for the ANN model (23). The model consisted of three individually trained sub-networks. An in-house developed Genetic Algorithm software module performed the initial training and network design selection, which resulted in a pool of 20 to 30 candidate networks. The remaining portion of the training was then completed using the conjugate gradient back-propagation learning algorithm. The selection of the final sub-network was based on the in-training validation results.

Statistical Analysis

In descriptive statistics, Mann-Whitney U test was used to assess difference in distributions of tumor markers or the ANN-derived composite index between different patient populations. To compare the diagnostic power of the ANN-derived composite index and CA125II alone, analysis and comparison of correlated receiver-operating-characteristic (ROC) curves was performed using the software package ROCKIT (Department of Radiology, the University of Chicago). The sensitivities were also compared between the composite index and CA125II by fixing specificity at a clinically relevant value.

RESULTS

The allocations of specimens to the three data sets are listed in Table 1. In the following, results on the in-training cross-validation set were not reported due to the small number of epithelial ovarian cancer cases and the lack of any early stage cases.

In Tables 2, descriptive statistics of the four tumor markers and the ANN results of the training set (Table 2A) and the independent test set (Table 2B) are listed. Statistically significant differences ($p < 0.000001$) were noted between epithelial ovarian cancer and apparently healthy women in the training set for all tumor markers and the ANN. In the test set, the relative elevations of the four markers among cancer patients, in either means or medians, were smaller than those observed in the training set, reflecting possible differences among the clinical sites. However, the differences for the four markers and the ANN remained statistically significant ($p = 0.000379 \sim < 0.000001$).

Figure 1 compares side by side the histograms of CA 125II and the ANN-derived composite index for women in the test set. In Figure 1A, when CA 125II was >55 U/ml, the number of early stage epithelial ovarian cancer patients rose abruptly. However, a large number of the cancer patients (54%) were distributed over the low CA125 range of 5-35 U/ml and overlapped with the healthy women. When the ANN was plotted in Figure 1B, overlap was also observed,

but the majority of patients with early stage ovarian cancer (71%) had an ANN value > 0.5 , a range with very few overlapped healthy women.

The added value of the ANN is visually demonstrated in the scatter-plot in Figure 2 where the ANN results of the independent test samples are plotted against their CA125II results. In the figure, a fair number of early stage ovarian cancer patients with a relatively low serum CA125II level had an elevated ANN result. For example, at a cutoff of 0.5 and a corresponding false positive rate of 2%, the ANN could detect five (including four invasive) early stage ovarian cancer patients who had a CA125II < 35 U/ml. Furthermore, the ANN would be able to detect 13 (including 12 invasive) early stage ovarian cancer patients who would otherwise be missed by CA125II alone if the cutoff of CA125II was to be set to match the 2% false positive rate.

In ROC curve analysis using only data from healthy donors and early stage epithelial ovarian cancer patients, the ANN-derived composite index had a higher overall diagnostic power. On the training data set, the estimated areas-under-curve (AUCs) for CA 125II and the ANN were 0.9356 and 0.9756, respectively ($p = 0.0429$, figure not shown). On the external independent test data set, the AUCs for CA125II and ANN were 0.9089 and 0.9449, respectively ($p = 0.0333$, Figure 3). ROC curves for the other markers (CA 72.4, M-CSF and CA 15-3) were also plotted in Figure 3. Since these markers individually do not perform as well as CA125II, relative improvements in AUC by the ANN index over them were more significant ($p = 0.0007 \sim < 0.0001$).

In the test data set, CA 125II exhibited a sensitivity of 65% for early stage ovarian cancer and 67% for invasive disease using a cut-off at 35 U/ml (Table 3), but there were 6% false positives. When the false positives were held constant at 2% for CA 125II and for the ANN, CA 125II detected only 46% of all early stage disease and 43% of invasive disease, whereas the ANN detected 71% of all early stage and 71% of invasive ovarian cancers. In Table 3, the separately estimated 95% confidence intervals of the sensitivities of CA125 and the ANN overlapped slightly. However, the differences in sensitivities between CA125 and the ANN were statistically significant in test for difference in paired proportions ($p = 0.047$ for all early stage epithelial cancers and $p = 0.040$ for invasive cases only).

DISCUSSION

Our analysis suggests that the combined use of multiple serum tumor markers through ANN improves both sensitivity and specificity for detection of Stage I epithelial ovarian cancer. In ROC curve analysis, ANN proved superior to the use of the best individual marker CA125II. At a fixed specificity of 98%, the differences in sensitivities between the ANN and CA125II were statistically significant for detecting all early stage cases and for detecting the invasive early stage cases only. The total number of early stage cases for independent test in the current study is still relatively small. Additional studies with larger early stage case populations may be needed to further strengthen the statistical power of the evidence. However, the differences, if established, would be clinically very significant. As shown by ROC curve comparison (figure 3), the present report supports the conclusion that ANN analysis of multiple markers can potentially permit improved sensitivity without sacrificing specificity when compared to the use of the individual markers alone.

Due to the relatively low prevalence yet high mortality of ovarian cancer, an effective screening strategy for ovarian cancer in general needs to attain minimally a positive predictive value of 10%, i.e. 10 operative procedures for each case of ovarian cancer detected. With a prevalence of 1 in 2,500 among postmenopausal women in the United States, this equates to requiring a test to have a sensitivity of 75% and a specificity above 99.7%. Even with the improved sensitivity and specificity, the current ANN algorithm using the four serum tumor markers by

itself will not meet this requirement. ANN analysis of multiple tumor markers does, however, appear promising as an initial step in a two-stage strategy where an abnormal blood test triggers a second test such as transvaginal sonography (TVS). In early clinical studies, a combination of CA 125 and transabdominal ultrasound produced a specificity of 99.7% (17,24). In general, in a two-stage test strategy, improvement in specificity (hence positive predictive value) is often at the expense of a lowered sensitivity (hence worse negative predictive value). A first-stage test with a much improved sensitivity while maintaining the same level of specificity will certainly make such a trade-off in the overall screening strategy a more viable choice. It needs to be noted, however, that the effectiveness of transvaginal sonography in differentiating malignant from benign masses could be affected by access to quality ultrasound examination (e.g., equipment and experience of the operators) and patient conditions (e.g., menopausal status).

The objective of the current paper is to evaluate the performance of the ANN-derived index in detecting early stage epithelial ovarian cancer. The inclusion of late stage epithelial ovarian cancer in the training data set to increase the otherwise relatively small number of cancer cases was based on the assumption that at the molecular level, or at least as far as the four tumor markers were concerned, early and late stage ovarian cancers represent only the quantitative progression of the same disease. Consequently, the expression patterns among the multiple tumor markers captured by the ANN from the mixture of early and late stage patients may be applicable to the test data of healthy donors and early stage patients. A drawback of the current study is that specificity of the ANN was not assessed on independent test samples of patients with benign diseases. In a previous report, an ANN that was closely related to the current ANN (the two shared a significant portion of training samples and had a common design) had a specificity for benign cases in the 90% range (21).

The distributions of CA 125II among healthy donors and among early stage ovarian cancer patients are statistically significantly different ($p < 0.000001$). However, as indicated by the histogram plot in Figure 1A, CA 125II exhibits a distribution among healthy women with a low, yet long, tail that spreads across almost the entire distribution range of CA 125II in patients with ovarian cancer. As a result, the sensitivity of CA 125II deteriorates significantly when the cutoff is set for a high specificity. In Figure 1B, the overlap between the distributions of the ANN composite index in the two diagnostic groups is much smaller.

The longitudinal algorithm developed by Skates (16) assesses one's risk of having ovarian cancer by detecting the change-point in sequential tests of the single assay CA 125II (i.e., the point at which CA 125II changing from normal biological variations to disease induced elevation). It was reported based on preliminary data that its overall positive predictive value exceeded the minimum requirement for it to be a viable tool for ovarian cancer screening. However, ovarian cancer is a heterogeneous disease. It could still be difficult for a single marker, even measured over time, to achieve a level of robust performance to allow for cost-effective screening of a general asymptomatic population. Result from the current study indicates the potential of combining multiple markers to improve both sensitivity and specificity. It is conceivable that by incorporating existing or newly discovered tumor markers in multivariate predictive models, a similarly devised longitudinal algorithm could improve significantly the sensitivity and specificity in detecting true early signs of ovarian cancer.

Acknowledgements

This study was partially supported by an NCI Grant 1P50 CA83639, UTMDACC Specialized Programs of Research Excellence (SPORE) in Ovarian Cancer and a Department of Defense IDEA grant DAMD17-OC03-IDEA.

References

1. Berek, JS.; Thomas, GM.; Ozols, RF. Cancer Medicine. 5. 2000. Ovarian Cancer; p. 1687-1720.

2. Bourne TH, Campbell S, Reynolds KM, et al. Screening for early familial ovarian cancer with transvaginal ultrasonography and colour blood flow imaging. *Bmj* 1993;306(6884):1025–1029. [PubMed: 8490496]
3. van Nagell JR Jr, DePriest PD, Reedy MB, et al. The efficacy of transvaginal sonographic screening in asymptomatic women at risk for ovarian cancer. *Gynecol Oncol* 2000;77(3):350–356. [PubMed: 10831341]
4. Sato S, Yokoyama Y, Sakamoto T, Futagami M, Saito Y. Usefulness of mass screening for ovarian carcinoma using transvaginal ultrasonography. *Cancer* 2000;89(3):582–588. [PubMed: 10931457]
5. Yin BW, Dnistrian A, Lloyd KO. Ovarian cancer antigen CA125 is encoded by the MUC16 mucin gene. *Int J Cancer* 2002;98(5):737–740. [PubMed: 11920644]
6. Bast RC Jr, Klug TL, St John E, et al. A radioimmunoassay using a monoclonal antibody to monitor the course of epithelial ovarian cancer. *N Engl J Med* 1983;309(15):883–887. [PubMed: 6310399]
7. Bast RC Jr, Siegal FP, Runowicz C, et al. Elevation of serum CA 125 prior to diagnosis of an epithelial ovarian carcinoma. *Gynecol Oncol* 1985;22(1):115–120. [PubMed: 2410329]
8. Zurawski VR Jr, Knapp RC, Einhorn N, et al. An initial analysis of preoperative serum CA 125 levels in patients with early stage ovarian carcinoma. *Gynecol Oncol* 1988;30(1):7–14. [PubMed: 2452773]
9. Einhorn N, Bast RC Jr, Knapp RC, Tjernberg B, Zurawski VR Jr. Preoperative evaluation of serum CA 125 levels in patients with primary epithelial ovarian cancer. *Obstet Gynecol* 1986;67(3):414–416. [PubMed: 2418395]
10. Jacobs I, Bast RC Jr. The CA 125 tumour-associated antigen: a review of the literature. *Hum Reprod* 1989;4(1):1–12. [PubMed: 2651469]
11. Woolas RP, Xu FJ, Jacobs IJ, et al. Elevation of multiple serum markers in patients with stage I ovarian cancer. *J Natl Cancer Inst* 1993;85(21):1748–1751. [PubMed: 8411259]
12. van Haaften-Day C, Shen Y, Xu F, et al. OVX1, macrophage-colony stimulating factor, and CA-125-II as tumor markers for epithelial ovarian carcinoma: a critical appraisal. *Cancer* 2001;92(11):2837–2844. [PubMed: 11753957]
13. Jacobs IJ, Skates SJ, MacDonald N, et al. Screening for ovarian cancer: a pilot randomised controlled trial. *Lancet* 1999;353(9160):1207–1210. [PubMed: 10217079]
14. Bast RC Jr, Urban N, Shridhar V, et al. Early detection of ovarian cancer: promise and reality. *Cancer Treat Res* 2002;107:61–97. [PubMed: 11775462]
15. Woolas RP, Conaway MR, Xu F, et al. Combinations of multiple serum markers are superior to individual assays for discriminating malignant from benign pelvic masses. *Gynecol Oncol* 1995;59(1):111–116. [PubMed: 7557595]
16. Skates SJ, Xu FJ, Yu YH, et al. Toward an optimal algorithm for ovarian cancer screening with longitudinal tumor markers. *Cancer* 1995;76(10 Suppl):2004–2010. [PubMed: 8634992]
17. Rosenthal AN, Jacobs IJ. The role of CA 125 in screening for ovarian cancer. *Int J Biol Markers* 1998;13(4):216–220. [PubMed: 10228904]
18. Baxt WG, Skora J. Prospective validation of artificial neural network trained to identify acute myocardial infarction. *Lancet* 1996;347(8993):12–15. [PubMed: 8531540]
19. Bottaci L, Drew PJ, Hartley JE, et al. Artificial neural networks applied to outcome prediction for colorectal cancer patients in separate institutions. *Lancet* 1997;350(9076):469–472. [PubMed: 9274582]
20. Muzzonigro G, Galosi AB. Biological selection criteria for radical prostatectomy. *Ann N Y Acad Sci* 2002;963:204–212. [PubMed: 12095945]
21. Zhang Z, Barnhill SD, Zhang H, et al. Combination of multiple serum markers using an artificial neural network to improve specificity in discriminating malignant from benign pelvic masses. *Gynecol Oncol* 1999;73(1):56–61. [PubMed: 10094881]
22. Xu FJ, Ramakrishnan S, Daly L, et al. Increased serum levels of macrophage colony-stimulating factor in ovarian cancer. *Am J Obstet Gynecol* 1991;165(5 Pt 1):1356–1362. [PubMed: 1957862]
23. Bishop, CM. *Neural Networks for Pattern Recognition*. Oxford: Oxford University Press; 1995.
24. Jacobs IJ, Skates S, Davies AP, et al. Risk of diagnosis of ovarian cancer after raised serum CA 125 concentration: a prospective cohort study. *Bmj* 1996;313(7069):1355–1358. [PubMed: 8956699]

Figure 1.

Comparison of histograms of A) CA 125II and B) the ANN-derived composite index among healthy donors and early stage epithelial ovarian cancer patients in the independent test data.

Figure 2.

Scatter plot of the independent test data compares CA 125II against the ANN-derived composite index. Darkened lines indicate cutoff values of ANN at 0.5, CA125II at 35 U/ml, and CA125II with 2% false positive rate. Notice that a fair number of patients with early stage epithelial ovarian cancer yet a relatively low CA125II level had an elevated ANN index value. Note that some of the samples overlapped in the plot.

Figure 3.
ROC curves and AUCs of individual tumor markers and the ANN-derived composite index based in the independent test data.

Table 1

Allocation of 468 samples into three data sets for training, in-training validation, and testing. Other than the 10 stage I borderline tumor cases in the independent test set, all epithelial ovarian cancer cases are invasive cases.

	Training Set	In-Training Cross Validation Set	Independent Test Set	Column Total
Healthy Donors	100	27	98	225
Benign Conditions Arising from Ovary	45	56		101
Epithelial Ovarian Cancer	24 (stage I) 3 (stage II) 28 (late Stages)	8 (late stages)	38 (stage I) 4 (stage II) 10 (stage I borderline tumor)	115
Other Malignant Pelvic Masses		27		27
Row Total	200	118	150	468

Descriptive statistics of four individual tumor markers and the ANN-derived composite index for healthy donors and early stage epithelial ovarian cancer patients. Statistical significances were estimated using the Mann-Whitney U test. A) Training data (only those in the two diagnostic groups); and B) independent test data. Units for all four markers were U/ml, and for the ANN an arbitrary unit (a.u.).

Table 2

	Early Stage Epithelial Ovarian Cancer n=27					Apparently Healthy Women n=100					p-value
	Mean	Median	Min	Max	SD	Mean	Median	Min	Max	SD	
MCSF	4.0	3.4	0.5	13.9	2.9	1.6	1.4	0.1	4.0	0.7	<0.0000001
CA153	56.3	27.3	12.3	355.0	71.5	14.5	13.3	3.1	34.6	6.6	<0.0000001
CA724	68.6	4.1	0.8	543.8	149.6	1.4	1.1	0.1	12.7	1.5	<0.0000001
CA125II	171.6	75.7	8.5	500.0	178.6	13.9	9.8	0.5	81.3	14.2	<0.0000001
ANN	0.59	0.60	0.34	0.91	0.13	0.25	0.22	0.14	0.57	0.09	<0.0000001

	Early Stage Epithelial Ovarian Cancer n=52					Apparently Healthy Women n=98					p-value
	Mean	Median	Min	Max	SD	Mean	Median	Min	Max	SD	
MCSF	3.4	3.1	0.8	8.0	1.6	1.9	1.5	0.1	4.7	1.0	<0.0000001
CA153	27.2	20.0	0.7	132.9	27.4	13.5	12.5	0.0	39.8	7.7	0.000379
CA724	12.0	2.3	1.0	157.4	28.8	1.4	1.3	0.3	4.1	0.8	<0.0000001
CA125II	92.5	48.7	10.1	620.2	128.0	13.2	10.4	0.0	63.7	12.3	<0.0000001
ANN	0.50	0.52	0.07	0.75	0.16	0.13	0.07	0.01	0.51	0.15	<0.0000001

Table 3

Comparison of diagnostic performance CA 125II and the ANN-derived composite index using the independent test data.

Performance 95% Conf. Interval	True Positive (All early stages)	True Positive (Invasive cases only) [¶]	False Positive
CA125II at cutoff = 35 U/ml	65% (34/52) CI: 51-78%	67% (28/42) CI: 50-80%	6% (6/98) CI: 3-13%
CA125II at 2% False Positive Rate	46% (24/52) [*] CI: 32-60%	43% (18/42) ^{**} CI: 28-59%	2% (2/98) CI: 0-8%
ANN at 2% False Positive Rate	71% (37/52) [*] CI: 57-82%	71% (30/42) ^{**} CI: 55-84%	2% (2/98) CI: 0-8%

* p = 0.047 for difference in sensitivity between CA125 and ANN;

** p = 0.040 for difference in sensitivity between CA125 and ANN.

[¶] excluding borderline cases.