Sains Malaysiana 44(12)(2015): 1671–1676

Performance Comparison of Serum and Urine Biomarkers from Independent Samples for Ovarian Cancer Screening

(Perbandingan Prestasi Serum dan Penanda Biologi Urin daripada Sampel Bebas untuk Saringan Kanser Ovari)

HYE-JEONG SONG, KI-SEOK JEONG, JONG-DAE KIM, CHAN-YOUNG PARK & YU-SEOP KIM*

ABSTRACT

This study compares the diagnostic performance of urine and serum multiple biomarkers for early diagnosis of ovarian cancer. The sample population includes 119 benign and 101 ovarian cancer patients. The marker combinations used to compare performance include 16 markers whose concentration values were obtained using the Luminex assay. In order to identify an optimal marker combination that could classify ovarian cancer and benign patients, the area under the curve (AUC) is used to evaluate 2-, 3-, and 4-marker combinations and the classification is performed by using logistic regression. In the case of urine samples, the best AUC values are 87.89% for the 2 protein markers combination, 90.22% for the 3 markers combination, and 92.43% for the 4 marker combination. In contrast, the best AUC values for serum sample are 92.4% for the 2 marker combination, 93.63% for the 3 marker combination and 94.63% for the 4 marker combination. This study confirmed that combining multiple biomarkers could improve diagnostic accuracy. Even though the urine sample shows relatively lower performance than serum, urine could be utilized more widely for its simple usability.

Keywords: Biomarker; ovarian cancer; serum; urine

ABSTRAK

Kajian ini membandingkan prestasi diagnostik air kencing dan beberapa penanda biologi serum untuk diagnosis awal kanser ovari. Gabungan penanda digunakan untuk membandingkan prestasi 16 penanda dan nilai kepekatan diperoleh dengan menggunakan logam penguji Luminex. Untuk menentukan gabungan penanda optimum yang boleh mengenal pasti kanser ovari dan pesakit biasa, luas di bawah lengkungan (AUC) digunakan untuk menilai kombinasi penanda 2-, 3- dan 4-. Pengelasan dijalankan dengan menggunakan regresi logistik. Dalam kes sampel air kencing, nilai AUC yang terbaik adalah 87.89% untuk 2 kombinasi penanda protein, 90.22% untuk 3 kombinasi penanda dan 92.43% untuk 4 kombinasi penanda. Sebaliknya, nilai AUC yang terbaik untuk sampel serum adalah 92.4% untuk 2 kombinasi penanda, 93.63% untuk 3 kombinasi penanda dan 94.63% untuk 4 kombinasi penanda. Kajian ini mengesahkan bahawa gabungan beberapa penanda biologi boleh meningkatkan ketepatan diagnostik. Walaupun sampel air kencing menunjukkan prestasi yang agak rendah daripada serum, air kencing boleh digunakan dengan lebih meluas untuk kebolehgunaan yang mudah.

Kata kunci: Air kencing; kanser ovari; penanda biologi; serum

INTRODUCTION

Ovarian cancer is a malignant tumor with the highest incidence in women between the ages of 50 and 70 years. Annually, 1,000-1,200 new cases of ovarian cancer were diagnosed and it is the second most common gynecologic cancer after cervical cancer. For all ovarian cancers, the 5-year survival rate is 44% and that of early-diagnosed epithelial ovarian cancer is 60–90%; however, when diagnosed after stage 3, the 5-year survival rate drops to 40% (American Cancer Society) (Table 1).

Clinical diagnosis of ovarian cancer requires surgical biopsy; however, for lesions in which ovarian cancer is suspected before surgery, various clinical tests must be carried out to check the stage of progression and metastasis of cancer organs. At the first diagnosis stage, vaginal ultrasound to check the outward shape and size of the tumor and serum CA-125 testing having suitable sensitivity for ovarian cancer detection were conducted (Nolen et al. 2009). As an ovarian cancer diagnostic method, pelvic examination helps clinicians assess the size, appearance and mobility of the ovaries; if they are hard with an irregular surface, whether the tumor is suspicious for malignant ovarian cancer can be determined. However, the sensitivity of pelvic examination for the diagnosis of ovarian cancer is quite low at 30% (Nolen et al. 2009). Therefore, ultrasound examination, developed by technological advances after the 1970s, has been used for ovarian cancer screening. Among many ultrasound examinations, transvaginal ultrasound approaches organs in the abdominal cavity with good resolution to effectively visualize the size and shape of ovaries and to check the presence of ovarian tumors, but it was difficult to distinguish between normal and abnormal states with this technique.

TABLE 1. Relative 5-year ovarian cancer survival rates

Stage	I	IA	IB	IC	II	IIA	–
Relative 5-year survival rate	89%	94%	91%	80%	66%	76%	–
Stage	IIB	IIC	IID	IIIA	IIIB	IIIC	IV
Relative 5-year survival rate	67%	57%	34%	45%	39%	35%	18%

The CA-125 test assesses serum CA-125 glycoprotein concentrations was used to diagnose epithelial ovarian tumors. Of the biomarkers used for ovarian cancer detection, CA-125 is the most commonly clinically used biomarker and 85% of the patients with epithelial ovarian cancer show positive reaction to the marker, with 35 U/mL of the cut-off point for positive reaction. In the case of early ovarian cancer (stages 1 and 2), only 50-60% show positive reaction and for late-stage ovarian cancer (stages 3 and 4), more than 90% of patients show positive reaction (American Cancer Society). However, CA-125 may also increase during endometriosis, regular menstruation, ovarian cysts and pregnancy; therefore, this test alone cannot be used to confirm the diagnosis of the cancer (Nolen et al. 2009). An early stage ovarian cancer patient does not show specific observable symptoms; expensive and unnecessary surgical examinations could be required. Epithelial cancers, which account for over 90% of all ovarian cancers, are usually discovered later than stage 3; therefore, the need for a diagnostic examination that could discover ovarian cancer at early stages has become increasingly important (Nolen et al. 2009).

For diagnosis of early-stage cancer, a single marker cannot provide sufficient sensitivity and specificity; therefore, a diverse combination of biomarkers is needed for early diagnosis of ovarian cancer (Pepe et al. 2006; Yukovetsky et al. 2013). An examination method using biomarkers, involving methods using serum or urine, is a comparatively simpler and less expensive method than other diagnostic tests used for early diagnosis of cancer (Feng et al. 2006).

A biomarker refers to a marker that can accurately determine whether an organism is pathologically normal or abnormal and how much the organism reacts to a specific drug. Specifically, a biomarker can determine the pathologic status of a disease and predict the extent of an organism's reaction when a specific drug is used to treat the disease. An ideal tumor marker is a protein detected in fragments from the patient's urine and serum, which were not found in normal individuals (Chatterjee & Zetter 2005; Hellstrom et al. 2010; Moore et al. 2008). The United States Food and Drug Administration (FDA) with the goal of cancer diagnosis, have approved the use of tumor-related biomarkers. In 2007, the FDA published regulations and guidelines regarding the use of '*in vitro* diagnostic multivariate index assays' (IVDMIAs). The FDA defines an IVDMIA as a combined device that combines biomarker values to create a specific output consisting of a classification, score and index using analytic functions, with the overall goal of diagnosis, mitigation, treatment and prevention (Food and Drug Administration). Normally, we do not have enough biomarkers showing almost 100% specificity to t specific disease. To resolve this problem, the IVDMIA could improve the diagnostic accuracy by combining multiple biomarkers and by statistically analyzing the numbers related to the markers. Unlike a single biomarker presupposing only one single value, the IVDMIA combines multiple values from multiple biomarkers which complement each other. Therefore, it yields better results than biomarkers used individually (Zhang 2012).

Studies on biomarkers for diagnosis of ovarian cancers normally use serum biomarkers instead of urine. But urine biomarkers are easily handled clinically and are also used for a completely non-surgical tumor detection method that can detect cancer patients earlier among patients with positive reaction (Kim et al. 2010; Nolen & Lokshin 2012; Petri et al. 2010).

Petri et al. (2010) compared the receiver-operating characteristic (ROC) area under the curve (AUC) values between serum and urine samples in the same patient group, showing little ROC values different between the urine biomarker at 84% and the serum biomarker at 83%. The majority of the women with clinical signs of ovarian cancer have benign conditions; therefore, a method capable of differentiating between benign conditions and ovarian cancer would be beneficial (Amonkar et al. 2009).

The present study aimed to evaluate the diagnostic performance between urine and serum biomarkers, which classify the benign patients and the ovarian cancer patients. A total of 16 similar urine and serum biomarkers were used to construct 2-4-biomarker combinations to evaluate the AUC of the ROC and diagnostic performance of the optimal marker combination was assessed using logistic regression.

Data Collection

The sample population for urine biomarker evaluation consisted of 119 benign patients and 101 ovarian cancer patients and that for serum biomarker evaluation consisted of 119 benign and 101 ovarian cancer patients. Even though the numbers of urine and serum samples were the same, those samples are collected from different patients groups. This was due to the difficulty in simultaneously obtaining urine and serum biomarker samples from the same patient.

TABLE 2. Clinical samples for serum & urine biomarker testing

Serum		Urine	
Characteristics	No. of patients	Characteristics	No. of patients
No. of patients studied	220	No. of patients studied	220
Benign Tumor	119	Benign Tumor	119
Ovarian Cancer	101	Ovarian Cancer	101
Figo Stage		Figo Stage	
I	34	I	26
II	15	II	9
III	37	III	54
IV	15	IV	12

In this study, 16 different urine and blood biomarkers were used. A total of 220 Korean women samples from the ASAN Medical Center (2010) were included in the study. Tables 2 and 3 present information about the tested clinical samples.

The urine and serum biomarker concentrations were assessed by using the Luminex (Nolen & Lokshin 2012) multiplex immunoassay method and an ovarian tumor-specific biomarker immunoassay kit (Borgia et al. 2009). The analysis was conducted according to the protocol provided by Luminex and the samples were analyzed using the Bio-Plex suspension array system (Jung et al. 2009). The expression level of biomarkers was set as the medium fluorescence intensity produced from 50-100 microbeads per analyte per sample. The analyte concentration was assessed using Bio-Rad 5-parameter curve fitting at medium-level fluorescent intensity (Zhang 2012).

METHODS

In order to identify an accurate biomarker combination with a high diagnostic accuracy for ovarian cancer, the performance of urine and serum biomarker combinations was compared. To determine biomarker combinations that can best differentiate between the ovarian cancer and benign cases, the ability to differentiate between subsets of combined markers was assessed to find the optimal subset (Saeys et al. 2007). Sensitivity and specificity must both be evaluated to choose a model with appropriate differentiation performance. A standard method for harmonizing sensitivity and specificity includes calculating the AUC of the ROC (Pepe et al. 2006). This study used logistic regression to evaluate the AUC of the ROC and choose marker combinations with the highest values.

Logistic regression is a standard statistical algorithm used when samples are divided into 2 or more groups (multivariate data) and predicts a proper group in which individual observation will fall. The relationship between input patterns (independent variable) and resulting value (dependent variable) is quantified. If the probability of an event of independent variable is p, the odds value of this event is obtained and through logit transformation S curve is transformed linearly, showing the specific characteristic of the regression model.

$$p = \frac{1}{1-\exp\left(-a-\left(b_1x_1+b_2x_2+\ldots+b_nx_n\right)\right.}$$
$$odds: \frac{p}{1-p} = \exp\left(a+b_1x_1+b_2x_2+\ldots+b_nx_n\right). \quad (1)$$
$$Logit: \log\left(\frac{p}{1-p}\right) = a+b_1x_1+b_2x_2+\ldots+b_nx_n$$

In this case, x_n is the independent variable and b_n is the logistic regression value. The regression value is calculated using the Newton-Raphson method of repetitive analysis. Setting the value obtained from the regression model (1) as z, the probability P is calculated using (2).

$$p = \frac{1}{1+\exp^{-z}}. \quad (2)$$

The cut-off value is applied to this probability to determine into which type it will fall (Kohavi 1995).

The average AUC value was calculated by repeating 5-fold cross validation 1000 times to choose the best urine and serum biomarker combination. Lastly, the combination with the highest average AUC was chosen as the optimal biomarker combination. K-fold cross validation is a statistical analysis method used to test collected samples. The collected data samples were divided into k samples mutually exclusively. One sub-sample is left as validation data for the model test and the remaining k - 1 sub-sample were used as training data. Each time a sub-sample was used as validation data, k sub-samples are repeated k times during the cross validation process. Each k results from each step and component of this process were used to calculate an average to create a single evaluation assay used for the study (Kohavi 1995). Bootstrap estimates are able to evaluate the potential of a model only by using the training data. Within this method, the sample's hold-out testing independence can be maintained (Amonkar et al. 2009).

Figure 1 shows an average AUC convergence graph based on a 5-fold cross validation. The value changed rapidly until it was repeated approximately 300 times. The average AUC stabilized after the 1000th 5-fold cross validation.

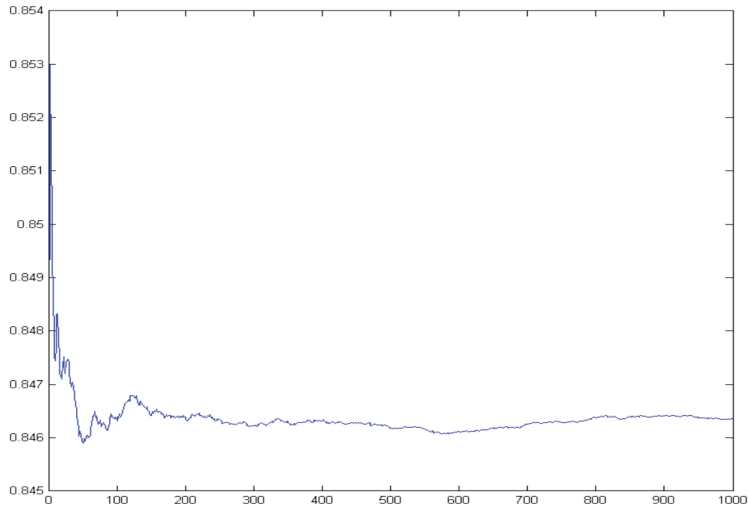


FIGURE 1. Average AUC convergence graph

RESULTS AND DISCUSSION

For assessment of urine and serum biomarker sample data, we used a logistic regression algorithm to calculate the AUC, sensitivity, specificity, distribution, accuracy, positive predictive value (PPV) and negative predictive value (NPV) using leave-one-out cross validation, without revealing the names of the biomarkers to avoid violation of patent laws. Table 3 shows the top 3 diagnostic performances of 2-urine-biomarker combinations and top 3 diagnostic performances of 2-serum-biomarker combinations (%).

Urine biomarker combination with the best performance was the marker combination M9 and M8 with an AUC of 89.89% and accuracy of 77.27%. Serum biomarker combination with the best performance was the combination M15 and M9 with an AUC of 92.40% and accuracy of 82.73%.

Figure 2(a) and 2(b) shows the ROC curve of a 2-biomarker combination of urine and serum markers with good performance.

Table 4 shows the top 3 diagnostic performances of 3-urine-biomarker combinations and top 3 diagnostic performances of 3-serum-biomarker combinations. The 3-urine-biomarker combination with the best performance was M13, M9, and M8 with an AUC value of 90.22% and accuracy of 82.27%. The 3-serum-biomarker combination with the best performance was M15, M13, and M9 with an AUC value of 93.63% and accuracy of 83.64%.

TABLE 3. Top 3 diagnostic performances of 2-urine & serum-biomarker combinations (%)

	Urine				Serum		
Markers	M9,M8	M13,M9	M9,M6	Markers	M15,M9	M13,M9	M1,M9
AUC	87.89	86.6	86.27	AUC	92.4	91.51	89.34
95% CI	85.78~92.14	80.09~91.11	80.52~90.06	95% CI	88.46~95.26	86.83~94.66	83.69~92.94
Sensitivity	57.43	60.4	57.43	Sensitivity	69.31	66.34	59.41
Specificity	94.12	98.32	96.64	Specificity	94.12	94.12	98.32
Accuracy	77.27	80.91	78.64	Accuracy	82.73	81.36	80.45

TABLE 4. Top 3 diagnostic performances of 3-urine&serum-biomarker combinations (%)

	Urine				Serum		
Markers	M13,M9 M8	M9,M7 M8	M9,M6 M8	Markers	M15,M13 M9	M7,M15 M9	M6,M15 M9
AUC	90.22	88.86	88.81	AUC	93.63	93.39	93.26
95% CI	86.41~94.74	84.06~93.03	84.06~92.99	95% CI	90.21~96.20	89.90~95.82	89.65~95.65
Sensitivity	62.38	63.37	59.41	Sensitivity	78.28	70.6	70.3
Specificity	99.16	94.96	96.64	Specificity	93.28	95.8	94.96
Accuracy	82.27	80.45	79.55	Accuracy	83.64	84.09	83.64

Figure 2(c) and 2(d) shows the ROC curves of the 3-urine-and-serum-biomarker combinations with good performance.

Table 5 shows the top 3 diagnostic performances of the 4-urine-biomarker combinations and top 5 diagnostic performances of the 4-serum-biomarker combinations. The 4-urine-biomarker combination with the best performance was M13, M9, M6, and M8 with an AUC of 92.43% and accuracy of 82.73%. The 4-serum-biomarker combination with the best performance was M6, M15, M13, and M9 with an AUC of 94.63% and accuracy of 85.91%.

Figure 2(e) and 2(f) shows the ROC curves of the 43-urine-and-serum-biomarker combinations with good performance.

Conclusion

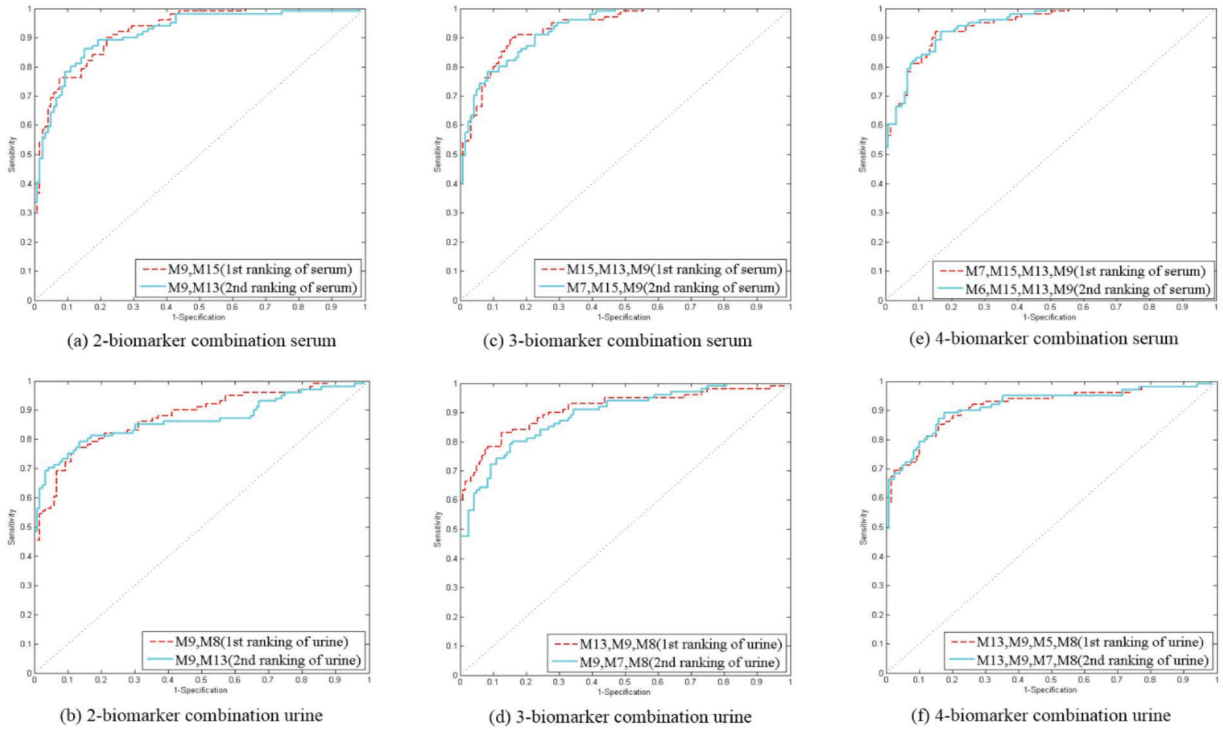
In this study, using serum and urine biomarker data, the optimal combination of 2, 3, and 4 biomarkers was used to evaluate their performance. Using logistic regression analysis, the score of each biomarker combination was calculated and 5-fold cross validation was repeated 1000 times to obtain the average AUC value. By choosing those with high average values, the AUC, sensitivity with 95% of specificity and accuracy were analyzed.

The performance of 2, 3 and 4 urine biomarkers were analyzed and those with the best performance included M9 and M8 at 87.89%; M13, M9, and M8 at 90.22%; and M13, M9, M6 and M8 at 92.43% for the AUC. This finding shows that using more biomarkers could lead to better performance.

The performance of 2, 3 and 4 serum biomarkers was analyzed and those with the best performance were M15 and M9 at 92.4%; M15, M13, and M9 at 93.63%; and M13, M6, M15 and M9 at 94.63% for the AUC. This finding also showed that using more biomarker combinations could lead to better performance. On comparing the effectiveness of urine and serum biomarkers, serum biomarkers were

TABLE 5. Top 3 diagnostic performances of 4-urine&serum-biomarker combinations (%)

	Urine				Serum		
Markers	M13,M9 M6,M8	M13,M9 M7,M8	M13,M9 M15,M8	Markers	M6,M15 M13,M9	M7,M15 M13,M9	M4,M15 M13,M9
AUC	92.43	91.56	91.47	AUC	94.63	94.37	94.26
95% CI	86.64~95.42	86.55~94.72	87.12~94.80	95% CI	91.10~96.65	90.41~96.60	90.65~96.44
Sensitivity	62.38	64.36	65.35	Sensitivity	77.23	75.25	72.28
Specificity	1	99.16	97.48	Specificity	93.28	93.28	93.28
Accuracy	82.73	83.18	82.73	Accuracy	85.91	85	83.64



(a) 2-biomarker combination serum (c) 3-biomarker combination serum (e) 4-biomarker combination serum

(b) 2-biomarker combination urine (d) 3-biomarker combination urine (f) 4-biomarker combination urine

FIGURE 2. Comparison of combinations of 2, 3, 4 urine and serum biomarkers

found to yield more accurate results than urine biomarkers, albeit to a small degree. However, normally, getting and handling urine sample are much easier and less expensive than serum sample. With this result, using urine sample rather than serum sample could be more recommendable for its practical issues even though it's very tiny lower AUC values than the serum.

In obtaining the optimal 4-biomarker combination, more time and cost were incurred. Therefore, further studies are needed to identify an optimal algorithm to obtain the most accurate biomarker combination in a cost- and time-efficient manner. The results of this study may aid in the early diagnosis of ovarian cancer.

ACKNOWLEDGEMENTS

This research was financially supported by the Ministry of Knowledge Economy (MKE), Korea Institute for Advancement of Technology (KIAT) and Gangwon Leading Industry Office through the Leading Industry Development for Economic Region and Bio-IT Research Center, funded by Ahn-Gook Pharmaceutical Co. Ltd.

REFERENCES

American Cancer Society. http://www.cancer.org/Cancer/OvarianCancer.

Amonkar, S.D., Bertenshaw, G.P., Chen, T.H., Bergstrom, K.J., Zhao, J., Seshaiah, P., Yip, P. & Mansfield, B.C. 2009. Development and preliminary evaluation of a multivariate index assay for ovarian cancer. *PLoS One* 4: 1-11.

ASAN Medical Center. 2010. http://eng.amc.seoul.kr/lang/AboutAMCController.do?forward=/jsp/en/AMCMain.jsp&sslpage=http%3A//medical.amc.seoul.kr/medservice/main/main.do.

Borgia, J.A., Basu, S., Faber, L.P., Kim, A.W., Coon, J.S., Kaiser-Walters, K.A., Fhied, C., Thomas, S., Rouhi, O., Warren, W.H., Bonomi, P. & Liptay, M.J. 2009. Establishment of a multi-analyte serum biomarker panel to identify lymph node metastases in non-small cell lung cancer. *Journal of Thoracic Oncology* 4(3): 338-347.

Chatterjee, S. & Zetter, B. 2005. Cancer biomarkers: Knowing the present and predicting the future. *Future Oncology* 1(1): 37-50.

Feng, Q., Yu, M. & Kiviat, N. 2006. Molecular biomarkers for cancer detection in blood and bodily fluid. *Critical Reviews in Clinical Laboratory Sciences* 43(5-4): 497-560.

Food and Drug Administration. 2007. Draft Guidance for Industry, Clinical Laboratories, and FDA Staff. *In vitro* Diagnostric Multivariate Index Assays. Rockville, MD: U.S. Department of Health and Human Services. pp. 1-15.

Hellstrom, I., Heagerty, P.J., Swisher, E., Liu, P., Jaffar, J., Agnew, K. & Hellstrom, K.E. 2010. Detection of the HE4 pretein in urine as a biomarker for ovarian neoplasms. *Cancer Letters* 296: 43-48.

Jung, S., Oh, E.J., Yang, C.W., Ahn, W.S., Kim, Y., Park, Y.J. & Han, K. 2009. Comparative evaluation of ELISA and Luminex panel reactive antibody assays for HLA alloantibody screening. *Korean Journal Laboratory Medicine* 29: 473-480.

Kim, Y., Koo, I., Jung, B.H., Chung, B.C. & Lee, D. 2010. Multivariate classification of urine metabolome profiles for breast cancer diagnosis. *BMC Bioinformatics* 11: 1-9.

Kohavi, R. 1995. A study of cross-validation and bootstrap for accuracy estimation and model selection. *International Joint Conference on Artificial Intelligence* 14(2): 1137-1145.

Moore, R.G., Brown, A.K., Miller, M.C., Skates, S., Allard, W.J., Verch, T., Steinhoff, M., Messerlian, G., DiSilvesto, P., Granai, C.O. & Bast. R.C. Jr. 2008. The use of multiple novel tumor biomarkers for the detection of ovarian carcinoma in patients with a pelvic mass. *Gynecol. Oncology* 108: 402-408.

Nolen, B., Marrangoni, A., Velikokhatnaya, L., Prosser, D., Winans, M., Gorelik, E. & Lokshin, A. 2009. A serum based analysis of ovarian epithelial tumorigenesis. *Gynecologic Oncology* 112(1): 47-54.

Nolen, B.M. & Lokshin, A.E. 2012. Multianalyte assay systems in the differential diagnosis of ovarian cancer. *Expert Opinion on Medical Diagnostics* 6(2): 131-138.

Pepe, M.S., Cai, T. & Longton, G. 2006. Combining predictors for classification using the area under the receiver operating characteristic curve. *Biometric* 62: 221-229.

Petri, A.L., Simonsen, A.H., Høgdall, E., Christensen, I.J., Kjaer, S.K., Yip, C., Risum, S., Pedersen, A.T., Hartwell, D., Fung, E.T. & Høgdall, C. 2010. Comparison of proteomic biomarker panels in urine and serum for ovarian cancer diagnosis. *Proteomics-Clinical Applications* 4(3): 304-314.

Saeys, Y., Inza, I. & Larranaga, P. 2007. A review of feature selection techniques in bioinformatics. *Bioinformatics* 23(19): 2507-2517.

Yukovetsky, Z., Skates, S., Lomakin, A., Nolen, B., Pulsipher, T., Modugno, F., Marks, J., Godwin, A., Gorelik, E., Jacobs, I., Menon, U., Lu, K., Badgwell, D., Bast, R.J. & Lokshin, A. 2013. Development of a multimarker assay for early detection of ovarian cancer. *Journal of Clinical Oncology* 28(13): 2159-2166.

Zhang, Z. 2012. An *in vitro* diagnostic multivariate index assay (IVDMIA) for ovarian cancer: Harvesting the power of multiple biomarkers. *Reviews in Obstetrics & Gynecology* 5(1): 35-41.

Hye-Jeong Song, Jong-Dae Kim, Chan-Young Park & Yu-Seop Kim
Department of Ubiquitous Computing
Hallym University, 1 Hallymdaehak-gil
Chuncheon, Gangwon-do 200-702
Korea

Ki-Seok Jeong
Department of Computer Engineering
Hallym University, 1 Hallymdaehak-gil
Chuncheon, Gangwon-do 200-702
Korea

Hye-Jeong Song, Ki-Seok Jeong, Jong-Dae Kim, Chan-Young Park & Yu-Seop Kim
Bio IT Research Center
Hallym University, 1 Hallymdaehak-gil
Chuncheon, Gangwon-do 200-702
Korea

*Corresponding author; email: yskim01@hallym.ac.kr

Received: 2 September 2014
Accepted: 23 June 2015