Iranian Journal of Mathematical Chemistry, Vol. 4, No.2, December 2013, pp. 163−175 **IJMC**

Feature selection and classification of microarray gene expression data of ovarian carcinoma patients using weighted voting support vector machine

S. MASOUM• AND S. GHAHERI

Department of Analytical Chemistry, Faculty of Chemistry, University of Kashan, I. R. Iran

(*Received October 1, 2012; Accepted Feb 18, 2013*)

ABSTRACT

We can reach by DNA microarray gene expression to such wealth of information with thousands of variables (genes). Analysis of this information can show genetic reasons of disease and tumor differences. In this study we try to reduce high-dimensional data by statistical method to select valuable genes with high impact as biomarkers and then classify ovarian tumor based on gene expression data of two patient groups. One group treated by standard therapies and survived, while another group didn't be cure and die after some times. In the first step we used weighted voting algorithm (WVA) for selecting impressive genes to reduce dimension, therefore eliminate noisy data and make analysis easier and then partial least square – discriminante analysis (PLS-DA) and support vector machine (SVM) methods have been applied for classification of diminished data. Results show that classification by PLS-DA can distinguish two groups somewhat but SVM is more efficient and sufficient classification method.

Keywords: weighted voting algorithm, support vector machine, tumor classification, ovarian cancer, gene expression data.

1. INTRODUCTION

In the past decades, chemometrics methods generally have been used to solve chemical problems. But today, there is an approach for using these methods to analysis of gene expression data [1-3]. DNA microarrays are capable of detecting the expression levels of thousand genes over a few tens of different samples simultaneously [4]. Because of such huge volume of data, there is an increasing attention in data mining field and extraction of precious and helpful information from a huge collection of data [5]. Using statistical method and data mining is necessary to understand the mechanism and process of human deceases. Human demise as cancer and tumor are because of gene expression changes, so

•Corresponding author (Email: *masoum@kashanu.ac.ir*)

discrimination between different group of samples or between patients and healthy people based on their gene expression is important step to take information out. By classification of individual gene expression, we have a model that can predict the class of a goal sample with unknown class label. The enormous amount of data and small size of individuals is a challenge of this kind of studies, because in such immense data, arrangement is complicated. Also this kind of data usually contains unsuitable and redundancy information [6]. Furthermore, before classification, reduce of dimension is important task. By decreasing dimension and select some genes through whole, we can use selected genes as biomarker to forecast deceases. Researchers introduce many methods for classification and selecting best features. For example, Sarhan developed a method based on artificial neutral network (ANN) and discrete cosine transforms (DCT) [7]. Zheng et al. proposed independent component analysis (ICA) method coupled by sequential floating forward selection (SFFS) technique [8]. Literatures also referred to PLS-DA frequently [9-11]. Furey et al. offered SVM method for classification of gene expression [2]. Other methods that have been used are radial basis function neural network [12] logistic discrimination and quadratic discriminante analysis [13].

Ovarian carcinoma is one of the most common type of gynecological cancers, is fifth reason of cancer demise in women [14]. Standard treatment in ovarian cancer patient is surgery followed by chemotherapy that some patient will be cured while others relapse. If these different patient groups could be identified before therapy, the alternative treatments or strategies might be used instead of standard treatments [15].

The data set have been used in this study is microarray analysis results of ovarian cancer patients. So, studied samples are consisting of gene expressions of two groups. One group didn't be cure and die after some time and another one are survived after 5 years. All samples are belonging to patients in stage III ovarian adecarcinoma. We used WVA to reduce dimension and then classify reduced data by PLS-DA and SVM methods.

2. METHODOLOGY

1.1. Weighted voting algorithm (WVA)

The first step in data analysis is valuation of genes as variables. Selecting genes with higher and lower expression is important task. Using these genes as prognostic factors may make easy identification of patients who are expected to relapse and die of the decease [15]. Furthermore, because of the large number of variables, recourse to conventional classification methods may be hard both for analytical and interpretive reasons. In this study we used WVA to select genes with higher and lower expression as biomarkers [16-17]. This algorithm calculates S_x value for each genes of data set according to equation 1.

$$S_x = \frac{(\mu_{c_1} - \mu_{c_2})}{(\sigma_{c_1} + \sigma_{c_2})} \quad (1)$$

S_x= weighted voting value for every gene

μ_c = mean of expression values in class 1 and class 2

σ_c = standard deviation of expression values in class 1 and class 2

The S_x value show how much is correlation of every genes with a particular distinction. Also it detects genes which have higher variance in one group but low variance in another one. This bias is useful for biological sample. For example, in cancer research, genes in normal tissue work normally and the regulation of which are strict. However, in tumors, genes are deregulated and levels of microarray data expressions vary widely [18]. After dimension reduction, the selected genes were used for classification.

1.2. Partial Least Square – Discriminant Analysis (PLS-DA)

PLS–DA, a special form of partial least square (PLS) modeling, aims to find the variables and direction in multivariate space, which discriminate the known class in training set. In PLS-DA, an indicator Y matrix of category variables is constructed which contains as many columns as there are known class in the training set. In this context, PLS-DA accomplishes a rotation of the projection to latent variable focusing on class separation [19]. PLS-DA score plot show distribution of two classes and root mean square error (RMSE) value reveals validity of separation. RMSE as prediction error parameter defined as below:

$$\text{RMSE} = \sqrt{\frac{\sum_i (Y_i - \hat{Y}_i)^2}{n}} \quad (2)$$

Y_i = real class for i_{th} sample

$\hat{Y}_i$ = predicted class for i_{th} sample

n= number of sample

It is obvious that the best parameter for RMSE is 0 when model can predict all classes exactly right.

3.2. Support Vector Machine (SVM)

The SVM algorithm originally introduced by V. Vapkin in 1998 [20]. For the first time, Furey et al. offered SVM method for microarray expression classification [2]. SVM classification is based on hyper-plan or a set of hyper-plans that separate labeled training data considering their classes so that the distance between them will be maximized. If in a

finite dimension space (linear), separation isn't possible, a much higher dimension on infinite space is used in combination with kernel techniques such as linear, polynomial, Gaussian radial basis and exponential radial basis function. It's clear that the hyper-plan which can classify the two classes of samples suitably isn't unique. To finding the best hyper-plan, called optimal separating hyper–plan (OSH), the concepts of margin, is introduced as distance of hyper-plan to nearest data point of each class (support vectors). There are many classifiers called hyper-plan that can separate the data, but there is only one that maximizes the margin [21]. Suppose problem of separating the set of training vectors belonging to two separate classes,

$$D = \{(x^1, y^1), \ldots, (x^m, y^m)\}, x \in R^n, y \in (1,-1) \quad (3)$$

The hyper–plan is:

$$\langle w, x\rangle + b = 0 \quad (4)$$

Where w is the normalized weight vector with the same dimension as x and b is the normalized bias of the hyper-plane, any hyper-plane f(x) should meet the following state:

$$\langle w, x\rangle + b \geq 1 \text{ for } y_i = 1 \quad (5)$$

$$\langle w, x\rangle + b \leq -1 \text{ for } y_i = -1 \quad (6)$$

So:

$$y_i (\langle w, x\rangle + b) \geq 1 \quad (7)$$

Then, the margin between the two paralleled hyper-planes can be written as:

$$Margin = \frac{2}{\|w\|} \quad (8)$$

Therefore, the structure of OSH can be transformed to the following optimizing problem:

Maximize: $\frac{2}{\|w\|}$

Subject to: $\langle w, x\rangle + b \geq 1$

By solving problem and finding OSH, we can classify a new data sample s. A label is assigned in according to its relationship to the decision boundary, and the corresponding decision function is:

$$f(s) = \text{sign}(\langle w,s\rangle + b) \quad (9)$$

4.2. Dataset

The data base resource currently available on the World Wide Web: www.ncbi.nih.gov/geo was a table (X-matrix) in which 56 individuals with 30000 probe sets (variables) reported. Each probe set contains one gene and it is also possible that one gene occupies more than one probe set. Individuals are belonging to two groups, 5–year survivor (class 1) and dead (class 2) patients. If consider X as a descriptor matrix, an appropriately selected dependant

variable matrix (the dummy matrix Y) designating membership to given class. The data set was divided into two sets of training (38 samples) and monitoring (16 samples). The training set was used to develop the model. Together with the performance of the training set, the performance of an independent set must also monitored (monitoring set) to obstruct the overtraining phenomena.

All computations and chemometrics analyses were executed with programs in Matlab v. 7. (The mathworks, Inc., Natick, MA, USA). Different algorithm has been proposed in the literature to perform SVM for classification [22-24]. The Lin's Lib SVM v. 2.33 algorithm was used in the present work [24].

2. RESULTS AND DISCUSSION

The weighted voting algorithm makes a weighted linear combination of relevant marker or informative feature obtained in the data set. Fifty probe sets with lowest and fifty probe sets with highest value of S_x were selected as biomarkers are listed in table 1 and 2, respectively. To evaluate the robustness of these biomarkers, the final step is to classify the data set. There have been many methods for performing the classification task. We used PLS-DA and SVM which have been proved to be very useful and robust to classify the microarray gene expression data.

Modeling by PLS-DA method was done on diminished training set. Validation of model was checked by monitoring set. Different preprocessing may help better discrimination. RMSE values with different preprocessing methods are arranged in table 3. In Figure 1, the result for the three latent variable normalize-PLS-DA model is shown. This figure shows that PLS-DA method can separate two groups somewhat but not completely. PLS-DA result based on original data set without any dimensional reduction in table 4 indicates that RMSE value for monitoring set is not satisfactory compare to reduce one.

Among different supervised methods, SVM seems to be the most suitable one, because for the classification purpose only support vectors are needed. This means that for the classification a limited number of data points are used and therefore the calculation processes would be reduced. In the present work, among 38 samples of the training set only a total of 14 samples were chosen as support vectors. When it is used for classification, SVM can separate a given set of binary labeled training data with a hyper-plan that is maximally distant from them (the maximal margin hyper-plan). For the case in which no linear separation is possible, they can work in combination with the technique of kernels, which automatically realize a nonlinear mapping to a feature space. Generally, the hyper-plan founded by SVM in a feature space corresponds to a nonlinear decision boundary in the original space. Linear SVM results show the 100% accuracy on training and 93% accuracy on monitoring set. Applying WVA-SVM on DNA microarray data can be considered as a powerful tool for tumor classification from gene expression data.

Previous study on such data set show that three genes are candidate biomarkers: TACC1 (transforming acidic coiled-coil containing protein 1), MUC5B (mucin 5 subtype B) and PRAME (preferentially expressed antigen in melanoma) [15]. The typical function of TACC1 is not accurately known, but observations have shown that the protein is concentrated at centrosomes during mitosis and may play a role in cytokinesis [25,26]. This gene known as a cancer related feature in literature [27-29]. MUC5B belongs to the mucin family of high-molecular-weight glycoproteins found in human epithelial cells. MUC5B, a secreted gel forming mucin, has been studied and play important role in a number of tumor types like breast and gastric cancer [30,31]. The function of PRAME in normal tissue is still unknown, but it encodes an antigen recognized by autologous cytolytic T lymphocytes and its expression is absent or low in normal adult tissue, except male germ cells [32].The effect of this gene, as cancer related gene in ovarian cancer has been verified in some researches [34,35]. Also abnormal expression of this gene is observed in melanoma and neuroblastoma cancer [35,36]. Results show S_x values for these three genes obtained by weighted voting algorithm are approximately in good agreement with other studies. When gene expression of survivor subgroup compared with remaining tumor, TACC1 and MUC5B are between highest S_x and PRAME is one of fifty genes with lowest S_x. Various genes with unknown function among the “top 100” (50 in table 1 and 50 in table 2) deserve high priority in future studies, that provide shortcuts in genome-based ovarian cancer research.

3. CONCLUSION

In this research, we presented WVA and SVM for feature selection and classification of tumor, based on microarray gene expression data. The methodological involve dimension reduction of high-dimensional gene expression data, followed by feature selection using WVA and classification by applying SVM. The results show that our method is effective and efficient in classifying ovarian tumor.

ACKNOWLEDGMENT

The authors are grateful to University of Kashan for supporting this work by Grant NO. 159181/3.

Table 1: Fifty Probe Sets with Lowest S_x.

No.	Identifier	S_x	No.	Identifier	S_x
1	ALPL	-0.8387509	26	EGFL6	-0.5806355
2	TSPYL5	-0.7305065	27	C22orf28	-0.5791292
3	AK023883	-0.7115028	28	LUC7L3	-0.5784992
4	Operon oligo ID: 300001540	-0.6912204	29	PRAME	-0.5778165
5	C10orf46	-0.6911821	30	KLHL24	-0.5755147
6	GPR137C	-0.6905153	31	RPN2	-0.5733304
7	ZNF250	-0.688837	32	ARID4B	-0.5729077
8	HSD17B14	-0.6707914	33	CST3	-0.5720629
9	TMTC1	-0.6657545	34	MAPK8IP1	-0.5713276
10	GALNT2	-0.6639226	35	HM13	-0.5698085
11	NASP	-0.6602168	36	RBM42	-0.5633995
12	XM_499130	-0.6564458	37	AK025101	-0.5625916
13	NRBP1	-0.6536662	38	HIST2H4A	-0.5591982
14	ASL	-0.6490638	39	SERTAD3	-0.5591699
15	TMTC1	-0.6417303	40	SEC61A1	-0.5590856
16	EFNA4	-0.6353448	41	FLJ21369	-0.555698
17	COL17A1	-0.6033936	42	ADAM17	-0.5539212
18	C19orf62	-0.6029183	43	CGN	-0.5528546
19	RNF185	-0.599269	44	SNRPB	-0.5528253
20	FAM84A	-0.5991216	45	COPB2	-0.5525224
21	DNMT3A	-0.5969614	46	ERH	-0.5511848
22	GPAA1	-0.5966022	47	IFIT1	-0.5480109
23	RBMY1J	-0.5936286	48	NOTCH3	-0.5475582
24	ATP2B4	-0.5850159	49	C20orf117	-0.5474255
25	C20orf117	-0.5840006	50	KRT6A	-0.544361

Table 2: Fifty Probe Sets with Highest S_x.

No.	Identifier	S_x	No.	Identifier	S_x
1	INTS10	0.77754612	26	SLC1A1	0.55741585
2	XM_036708	0.6892122	27	KIAA0415	0.55542346
3	Operon oligo ID: 300002269	0.68419241	28	BC012900	0.55444906
4	CDH3	0.67859078	29	POLR3D	0.55138735
5	TACC1	0.65699497	30	TTC18	0.55021117
6	SPATA2L	0.65636689	31	HIC2	0.54892664
7	Operon oligo ID: 200020491	0.6374128	32	CAPS	0.54833025
8	APOH	0.63583484	33	ABHD14B	0.54677712
9	FBXL21	0.63469991	34	DNAH9	0.54652939
10	GBE1	0.63243051	35	HCN2	0.54024783
11	MLC1	0.61475509	36	XM_496691	0.53942265
12	PSD	0.611211	37	NUDT6	0.53650802
13	CLU	0.61094734	38	Operon oligo ID: 300002652	0.53470338
14	XM_379145	0.60190264	39	ANKRD18A	0.53420268
15	WDR46	0.58562923	40	FYN	0.53414623
16	SAMD11	0.580747	41	FAM174A	0.53145332
17	MUC5B	0.57932012	42	UCN	0.53117405
18	PHLDA1	0.57808495	43	GIPC3	0.53083502
19	KIAA1462	0.57554463	44	C7orf34	0.53059775
20	RELN	0.56497197	45	CHD9	0.52966783
21	MRM1	0.56484086	46	NP_689472	0.52822182
22	Operon oligo ID: 200006958	0.56451803	47	LCAT	0.5280964
23	XM_496984	0.56330992	48	C16orf45	0.5277279
24	TSEN2	0.55978275	49	B9D1	0.52739697
25	NM_031306	0.55795636	50	KLHDC7A	0.52606313

Table 3: PLS-DA results with different preprocessing on reduced data.

	class 1	class 2	class 1	Class 2
Noun	0.15	0.15	0.35	0.37
Standard Normal Variate (SNV)	0.08	0.12	0.32	0.34
Orthogonal Signal Correction (OSC)	0.13	0.10	0.35	0.39
Normalize	0.10	0.12	0.30	0.32

Table 4: PLS–DA Results with Different Preprocessing on Original Data.

	class 1	class 2	class 1	Class 2
Noun	0.12	0.11	0.51	0.56
Standard Normal Variate (SNV)	0.06	0.06	0.43	0.47
Orthogonal Signal Correction (OSC)	0.05	0.05	0.43	0.46
Normalize	0.07	0.08	0.47	0.53

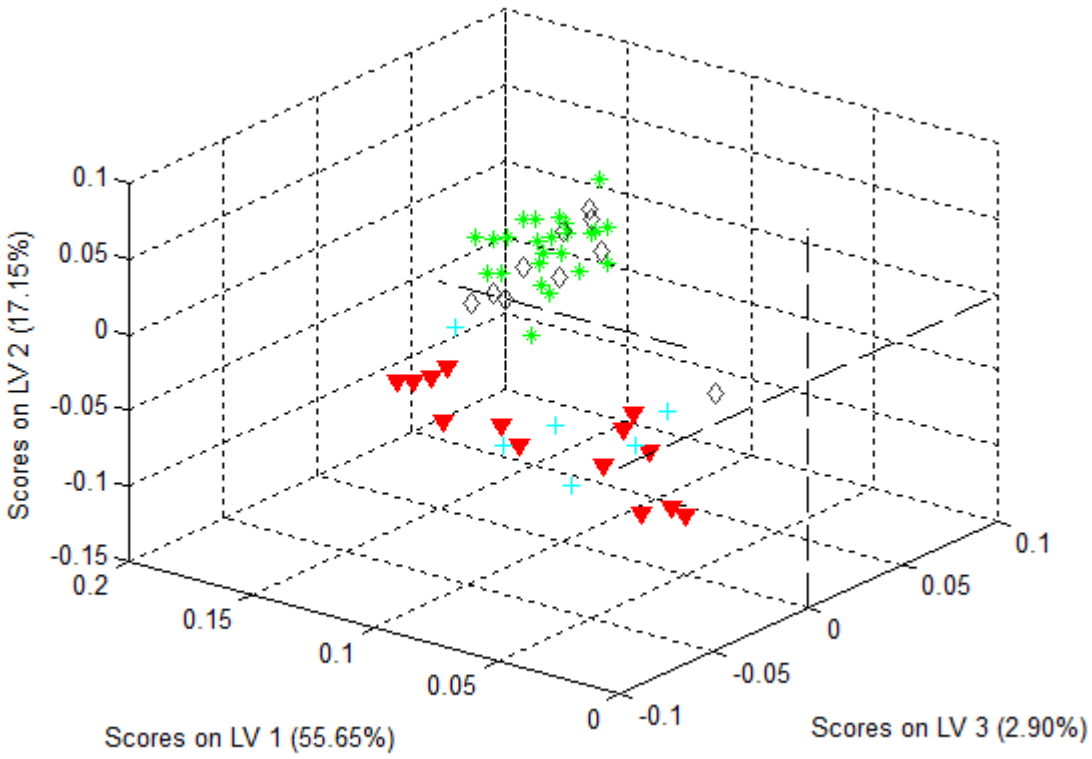

Figure.1: Score plot of the first three latent vectors. Training set: 5-year survivor (✱), dead (▼), corresponding monitoring set samples (◊,+).

REFERENCES

1. U. Alon and M. Barkai, Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays, P. Natl. Acad. Sci. USA. **96** (1999) 6745–6750.
2. T.S. Furey, N. Cristianini, N. Duffy, D.W. Bednarski, M. Schummer and D. Haussler, Support vector machine classification and validation of cancer tissue samples using microarray expression data, Bioinforma, **16** (2000) 906–914.

3. M. Bittner, P. Meltzer, Y. Chen, Y. Jiang, E. Seftor, M. Hendrix, M. Radmacher, R. Simon, Z. Yakhinik and N. Sampask, Molecular classi fcation of cutaneous malignant melanoma by gene expression profling, Nature. **406** (2000) 536–540.
4. C.H. Zheng, Y.W. Chong and H.Q. Wang, Gene selection using independent variable group analysis for tumor classification, Neural. Comput. Appl. **20** (2011) 161–170.
5. J.P. Bigus, Data mining with neural networks: solving business problems from application development to decision support, (McGraw-Hill, Hightstown New Jersey, 1996).
6. A. Osareh and B. Shadgar, Classification and diagnostic prediction of cancers using gene microarray data analysis, J. Appl. Sci. **9** (2009) 452–458.
7. A. Sarhan, Cancer Classification Based on Micro array Gene Expression Data Using DCT and ANN, J. Theor. Appl. Inf. Tech. **6** (2009) 208–216.
8. C. Zheng, D. Huang and L. Shang, Feature selection in independent component subspace for microarray data classification, Neurocomputing. **69** (2006) 2407–2410.
9. G. Musumarra, V. Barresi, D.F. Condorelli and S. Scirè, A bioinformatic approach to the identification of candidate genes for the development of new cancer diagnostics, Biol. Chem. **384** (2003) 321–327.
10. M. Pérez-Enciso and M. Tenenhaus, Prediction of clinical outcome with microarray data: A partial least squares discriminant analysis (PLS-DA) approach, Hum. Genet. **112** (2003) 581–592.
11. M. Barker and W. Rayens, Partial least squares for discrimination, J. Chemometr. **17** (2003) 166–173.
12. A. Castaño, F. Fernández-Navarro, C. Hervás-Martínez and P.A. Gutierrez, Neuro-logistic Models Based on Evolutionary Generalized Radial Basis Function for the Microarray Gene Expression Classification Problem, Neural. Process. Lett. **34** (2011) 117-131.
13. S. Dudoit, J. Fridlyand and T.P. Speed, Comparison of discrimination methods for the classification of tumors using gene expression data, J. Am. Stat. Assoc. **97** (2002) 77–87.
14. I.B. Runnebaum and E. Stickeler, Epidemiological and molecular aspects of ovarian cancer risk, J. Cancer. Res. Clin. **127** (2001) 73–79.
15. K. Partheen, K. Levan, L. Osterberg and G. Horvath, Expression analysis of stage III serous ovarian adenocarcinoma distinguishes a sub-group of survivors, Eur. J. Cancer, **42** (2006) 2846–2854.
16. S. Ramaswamy, K.N. Ross, E.S. Lander and T.R. Golub, A molecular signature of metastasis in primary solid tumors, Nat. Genet. **33** (2002) 49–54.

17. T. J. MacDonald, K.M. Brown, B. LaFleur, K. Peterson, C. Lawlor, Y. Chen, R.J. Packer, P. Cogen and D. A Stephan, Expression profiling of medulloblastoma: PDGFRA and the RAS/MAPK pathway as therapeutic targets for metastatic disease, Nat. Genet. **29** (2001) 143–152.
18. M. Reich, K. Ohm, M. Angelo, P. Tamayo and J.P. Mesirov, GeneCluster 2.0: an advanced toolset for bioarray analysis, Bioinforma. **20** (2004) 1797–1798.
19. K.P. Singh, A. Malik, D. Mohan, S. Sinha and V.K. Singh, Chemometric data analysis of pollutants in wastewater–a case study, Anal. Chim. Acta. **532** (2005) 15–25.
20. V. Vapnik, the nature of statistical learning theory, (Springer, New york, 1998).
21. http://www.eas.uccs.edu/wang/ECE5990/SVM.pdf, S R. Gunn, 1998.
22. http://www.csie.ntu.edu.tw/~cjlin/libsvm, C.-C. Chang and C.-J. Lin, 2002.
23. H. Li, Y. Liang and Q. Xu, Support vector machines and its applications in chemistry, Chemometr. Intell. Lab. **95** (2008) 188–198.
24. http://www.kernel-machines.org/
25. F. Gergely, C. Karlsson, I. Still, J. Cowell, J. Kilmartin and J.W. Raff, The TACC domain identifies a family of centrosomal proteins that can interact with microtubules, P. Natl. Acad. of Sci. USA. **97** (2000) 14352–14357.
26. B. Delaval, A. Ferrand, N. Conte, C. Larroque, D. Hernandez-Verdun, C. Prigent and D. Birnbaum, Aurora B -TACC1 protein complex in cytokinesis, Oncogene. **23** (2004) 4516–4522.
27. L.W. Chu, P. Troncoso, D. a Johnston and J.C. Liang, Genetic markers useful for distinguishing between organ-confined and locally advanced prostate cancer, Gene. Chromosome. Canc. **36** (2003) 303–312.
28. K. Partheen, K. Levan, L. Osterberg, K. Helou, G. Horvath, Analysis of cytogenetic alterations in stage III serous ovarian adenocarcinoma reveals a heterogeneous group regarding survival, surgical outcome, and substage, Gene. Chromosome. Canc. **40** (2004) 342–348.
29. R. Anbazhagan, H. Fujii and E. Gabrielson, Allelic loss of chromosomal arm 8p in breast cancer progression, Am. J. Pathol. **152** (1998) 815–819.
30. N. Berois, M. Varangot, C. Sóñora, L. Zarantonelli, C. Pressa, R. Laviña, J.L. Rodríguez, F. Delgado, N. Porchet, J. P. Aubert and E. Osinaga, Detection of bone marrow-disseminated breast cancer cells using an RT-PCR assay of MUC5B mRNA, Int. J. Cancer. **103** (2003) 550–555.
31. M. Perrais, P. Pigny, M.P. Buisine, N. Porchet, J.P. Aubert and I. Van Seuningen-Lempire, Aberrant expression of human mucin gene MUC5B in gastric carcinoma and cancer cells. Identification and regulation of a distal promoter, J. Biol. Chem. **276** (2001) 15386–15396.

32. H. Ikeda, B. Lethé, F. Lehmann, N. van Baren, J.F. Baurain, C. de Smet, H. Chambost, M. Vitale, A. Moretta, T. Boon and P.G. Coulie, Characterization of an antigen that is recognized on a melanoma showing partial HLA loss by CTL expressing an NK inhibitory receptor, Immunity. **6** (1997) 199–208.
33. T.R. Adib, S. Henderson, C. Perrett, D. Hewitt, D. Bourmpoulia, J. Ledermann, C. Boshoff, Predicting biomarkers for ovarian cancer using gene-expression microarrays, Brit. J. Cancer. **90** (2004) 686–692.
34. K. Hibbs, K.M. Skubitz, S.E. Pambuccian, R.C. Casey, K.M. Burleson, T.R. Oegema, J.J. Thiele, S.M. Grindle, R.L. Bliss and A.P.N. Skubitz, Differential gene expression in ovarian carcinoma: identification of potential biomarkers, Am. J. Pathol.**165** (2004) 397–414.
35. K.H. Lu, A.P. Patterson, L. Wang, R.T. Marquez, E.N. Atkinson, K.A. Baggerly, L.R. Ramoth, D.G. Rosen, J. Liu, I. Hellstrom, D. Smith, L. Hartmann, D. Fishman, A. Berchuck, R. Schmandt, R. Whitaker, D.M. Gershenson, G.B. Mills, R.C. Bast and N. Carolina, Selection of Potential Markers for Epithelial Ovarian Cancer with Gene Expression Arrays and Recursive Descent Partition Analysis, Anal. Clin. Cancer. **10** (2004) 3291–3200.
36. J.B. Welsh, P.P. Zarrinkar, L.M. Sapinoso, S.G. Kern, C. a Behling, B.J. Monk, D.J. Lockhart, R. A. Burger, and G.M. Hampton, Analysis of gene expression profiles in normal and neoplastic ovarian tissue samples identifies candidate molecular markers of epithelial ovarian cancer, P. Natl. Acad. of Sci. USA. **98** (2001) 1176–1181.