BIOINFORMATION
Discovery at the interface of physical and biological sciences

open access www.bioinformation.net Hypothesis

Volume 7(4)

An improved hybrid of SVM and SCAD for pathway analysis

Muhammad Faiz Misman*, Mohd Saberi Mohamad, Safaai Deris, Afnizanfaizal Abdullah, Siti Zaiton Mohd Hashim

Faculty of Computer Science and Information Systems, Universiti Teknologi Malaysia, 81310, Skudai, Johor Darul Takzim, Malaysia; Muhammad Faiz Misman – Email: faizmisman@gmail.com; *Corresponding author

Received September 29, 2011; Accepted October 02, 2011; Published October 14, 2011

Abstract:
Pathway analysis has lead to a new era in genomic research by providing further biological process information compared to traditional single gene analysis. Beside the advantage, pathway analysis provides some challenges to the researchers, one of which is the quality of pathway data itself. The pathway data usually defined from biological context free, when it comes to a specific biological context (e.g. lung cancer disease), typically only several genes within pathways are responsible for the corresponding cellular process. It also can be that some pathways may be included with uninformative genes or perhaps informative genes were excluded. Moreover, many algorithms in pathway analysis neglect these limitations by treating all the genes within pathways as significant. In previous study, a hybrid of support vector machines and smoothly clipped absolute deviation with groups-specific tuning parameters (gSVM-SCAD) was proposed in order to identify and select the informative genes before the pathway evaluation process. However, gSVM-SCAD had showed a limitation in terms of the performance of classification accuracy. In order to deal with this limitation, we made an enhancement to the tuning parameter method for gSVM-SCAD by applying the B-Type generalized approximate cross validation (BGACV). Experimental analyses using one simulated data and two gene expression data have shown that the proposed method obtains significant results in identifying biologically significant genes and pathways, and in classification accuracy.

Keywords: pathway analysis, smoothly clipped absolute deviation, support vector machines.

Background:
Incorporation of prior pathway data into microarray analysis has become a popular research area in bioinformatics due to the advantages in providing the further biological interpretation compared to the single gene microarray analysis. These advantages have triggered various experiments and approaches in order to identify informative genes and pathways that contribute to certain cellular processes. Two most common approaches to identify the informative genes and pathways are enrichment analysis (EA) and machine learning (ML) approaches **[1]**. In EA approaches, genes are viewed as functional pathways where the pathways with large number of differentially expressed genes between two conditions are considered as significant **[1]**. While in ML approaches, the genes in the pathways are discriminated between two or more phenotypes of interests and the pathway with genes that improves the prediction of the phenotype considered as significant pathways.

However, there are some challenges in pathway analysis, such as, some pathways may be included with uninformative genes or perhaps informative genes were excluded **[1]**. Another concern is that usually pathway data are curated from the biological context free; it can be that only several genes take part in certain cellular process when it goes to the phenotype specific analysis (e.g. lung cancer research). In order to deal with the challenges in the quality of pathway data, several researchers in EA approaches attempt to refining the pathway

data to specific conditions by removing the unaltered genes in the pathways, and including the additional information of the functional interpretation of the pathways or gene groups **[1].** In ML approaches, several researchers have included the gene selection methods in order to select the informative genes within pathways before the pathway evaluation process **[2, 3].** This is because, gene selection methods provide several advantages such as: (1) improves the classification accuracy, (2) removes uninformative genes, and (3) reduces computational time **[13].**

Misman *et al.* **[4]** had proposed the gSVM-SCAD in order to identify the significant genes and pathways, simultaneously dealing with the challenges in pathway data quality. gSVM-SCAD is a hybrid method of support vector machines (SVM) and smoothly clipped absolute deviation (SCAD) with group-specific tuning parameters that select the informative genes within pathways before the pathway evaluation process. gSVM-SCAD had shown its superior performance in identifying significant genes and pathways compared to the hybrid of SVM and L1 penalty function (L1 SVM), and almost with all other 5 classifiers without gene selection methods **[4].** However, in this work, we detected that gSVM-SCAD had some limitations.

A SCAD performance relies on the tuning parameter. If the tuning parameter is too small, it can bring little sparsity and over fit to the classifier model. While if it is too large, it can make very sparse to the classifier model but produced poor discriminating power **[5, 6].** Therefore, it is important to choose an appropriate tuning parameter selector method for the gSVM-SCAD, since the SCAD plays an important role in determining the true classification model for the SVM. The generalized cross validation (GCV) proposed by Fan and Li. **[5]** is widely used in the past literatures as a tuning parameter selector method for SVM-SCAD. Unfortunately, GCV had some limitations where it poorly performs when dealing with the low number of variables (genes) and large sample sizes **[6].** This is due to the reason that usually some pathways contained not more than 100 genes and even some pathways contained less than 10 genes. We believed that this scenario can lead to the poor performance of SCAD in selecting the informative genes and simultaneously identifying significant genes. In order to surmount the limitation of the gSVM-SCAD, we propose the B-type generalized approximate cross validation (BGACV) as a new tuning parameter selector method for gSVM-SCAD. Our proposed method is denoted as gSVM-SCADBGACV.

Methodology:

Experimental Data:

There are two types of data used: gene expression and pathway data. For gene expression data, the canine and lung cancer datasets were used to evaluate the performance of the gSVM-SCADBGACV (the information of the datasets is shown in **Table 1, see supplementary material**). For the pathway data, there are total of 441 pathways with 129 taken from KEGG database while other 312 pathways curated from BioCarta database. Both gene expression and pathway data can be downloaded at http://bioinformatics.med.yale.edu/pathway-analysis/datasets.htm. For the simulated data, total 1060 simulated genes are generated, where 848 genes are informative while other 212 genes are uninformative. The sample size for this experiment has been setup as 80. All 1060 genes are distributed into the Gi groups of genes where i = 1,...,15 and each group have a different size of genes. For the distribution of genes, each Gi contains i x 10 of genes. Moreover, 80% of genes in each group are informative. This creates 15 groups of genes with sizes varying from 10 to 150. The simulation data were created using the sim.data function in R package penalized SVM **[14].**

gSVM-SCAD with BGACV Tuning Parameter Method (The Proposed Method):

Our proposed method (gSVM-SCADBGACV) consists of three steps as shown in **Figure 1**. Step 1: group genes based on the pathway information provided by the pathway data. Step 2 is the most important step where the genes within pathways are evaluated using the SVM-SCAD, the uninformative genes are discarded from their pathway. In step 2.1, the proposed BGACV is used in order to select an appropriate tuning parameter for SVM-SCAD. In step 3, the classification error from the selected genes for each pathway is calculated. Step 1 and 3 is similar to the gSVM-SCAD but in step 2.1, there are some changes where BGACV is used instead of GACV.

Algorithm: gSVM-SCAD
Input: GE: Gene expression data
PD: Pathway data
λ : Tuning parameter
Output: SP: Significant pathways
IG: Informative genes
Begin
Step 1: Grouping genes based on their pathway information
For $j=1$ to max number of pathways in PD **do**
Find genes from GE that related to the pathway
Select and assign the related genes as a group
End-for
Step 2: Evaluate the pathways
For $j=1$ to max number of pathways in PD **do**
Step 2.1: Estimation of TP using a BGACV
For TP = 0.001 to 0.009 ,0.01 to 0.09 and 0.1 to 1 **do**
$BGACV(\lambda) = \frac{1}{n}\sum_{n=1}^{n}(1 - y_i\beta_{\lambda i}) + \frac{\log n}{2} DF_\lambda$
End-for
$\lambda = argmin_\lambda\{GACV(\lambda)\}$ // best λ produces minimum BGACV error
Step 2.2: Select the informative genes using the SVM-SCAD
Let β^k as the estimate of β at step k where $k = 0, \ldots, n$
The value of β^0 set by an SVM
While β^k not converge **do**
Minimizing the $\frac{1}{n}\Sigma[1-yf(x_i)]_+ + \Sigma_{j=1}^{d} P_\lambda(\beta_j)$
$k = k + 1$
If $\beta_j^k \le \epsilon$ **then** // $\epsilon = y_i - f(x_i)$
The gene j considered as non-informative and discarded
End-if
End-while
Step 2.3: Classify the selected genes using an SVM
Step 3: Calculate the classification error using a 10-fold cross validation
End-for
End

Figure 1: The gSVM-SCADBGACV procedure.

BGACV is a modified version of GACV where "B" stands for Bayesian information criterion (BIC) model selector proposed by Schwarz **[9].** Researchers believe that solution provided by BGACV is sparse and analogous to BIC **[6].** The first BIC type of GCV for SCAD had been proposed by Wang *et al.* **[10]** and improved by Wang *et al.* **[11]** to be more robust when dealing with the diverging number of parameters. Despite showing

good results when dealing with the diverging number of parameters, Wang *et al.* **[11]** admitted that their method still showed some limitation when dealing with the data with high number of samples and low number of variables. Therefore, we used the GACV formula **[12]** and transformed the GACV formula into BGACV. The GACV proved it robustness when dealing with data with high number of samples and low number of variables **[12].** The formula of GACV is given in **supplementary material**.

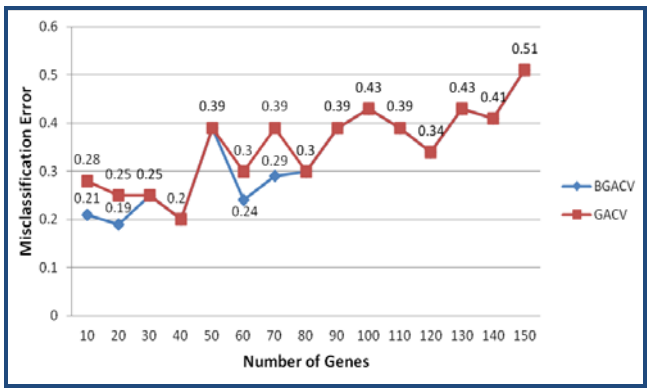

Figure 2: Correlations between the tuning parameter selector methods performance and the number of genes.

Discussion:

Simulation Study:

The purpose of this experiment is to evaluate the performance of both GACV and BGACV and providing the appropriate for gSVM-SCAD in a situation where the sizes of pathways are small and medium. Both gSVM-SCADBGACV and gSVM-SCAD were run for 10 times for each group of genes and the averages of the classification error are recorded. The results are presented in **Figure 2.**

From the **Figure 2,** it is shown that BGACV obtained less classification error compared to GACV in group of genes sizes 10, 20, and 60 with 0.02, 0.06, and 0.06 less, respectively. All these groups of genes contained a size of genes smaller than the sample size (sample size = 80). While for the groups of genes that contained the size of genes with equal and bigger than sample size, there are no different results between both BGACV and GACV methods. From the results above, it is shown that BGACV can deal with the small group of genes compared to GACV by producing less misclassification error. While for groups of genes that contained larger size of genes than the sample size, both BGACV and GACV methods have the same performance. Our analysis shows that BGACV are more robust and consistent in providing the near optimal for SCAD when dealing with the diverging numbers of parameters, especially when the number of genes are smaller than the number of samples.

Performance Evaluation:

A gSVM-SCADBGACV is also used on the lung cancer and canine datasets to evaluate the performance of the proposed method. From this experiment, results from the 10-fold cross validation (10-fold CV) were recorded and compared with other 5 supervised machine learning methods. The 5 supervised machine learning methods are PathwayRF **[15]**, neural networks, k-nearset neighbour with 1 neighbour (KNN1) and 3 neighbours (KKN3), and linear discriminant analysis (LDA). The results of the gSVM-SCADBGACV performance against other machine learning methods are shown in **Table 2 (see supplementary material)**.

For the canine dataset in **Table 2 (see supplementary material)**, it is interesting to note that gSVM-SCADBGACV outperforms gSVM-SCAD and the other machine learning methods with 85.95% accuracy. The gSVM-SCADBGACV achieved 4.87% higher than gSVM-SCAD, 3.62% higher than neural networks, slightly 0.04% higher than PathwayRF, and 4.96% higher than KNN1 and KNN3 respectively, and 9.34% higher than LDA. For the lung cancer dataset, gSVM-SCADBGACV also showed a significant improvement with 76.55% accuracy and 2.78% higher than gSVM-SCAD. While with other methods, gSVM-SCADBGACV obtained 4.55% higher than PathwayRF, 6.16% higher than neural networks, 14.82% higher than KNN1, 11.75% higher than KNN3, 13.31% higher than LDA, 21.41% higher than L1 SVM, 23.05% higher than SVM-SCAD.

From these results, it is proved that BGACV plays an important role in increasing the performance of SCAD in gSVM-SCAD. For the canine dataset, gSVM-SCAD showed quite poor performance, ranked third behind neural networks and PathwayRF. However, when BGACV is applied to the gSVM-SCAD, it showed a significant improvement even it obtained slightly higher than PathwayRF result. This is because, BGACV is consistently identified nearly optimal for SCAD and simultaneously provides nearly unbiased coefficient estimation when dealing with large coefficients. The results from the gSVM-SCADBGACV show significant achievement, since the pathway analysis usually dealing with the lower size of genes and larger size of samples.

Conclusion:

In this work, the proposed gSVM-CADBGACV is shown to outperform the gSVM-SCAD and other supervised machine learning methods. It is also shown that, a tuning parameter selector method plays an important role in gSVM-SCAD in dealing with the small genes size and large sample size that usually happened in pathway-based microarray analysis. The gSVM-SCADBGACV consistently showed better performance from the simulation study, the 10-fold CV accuracy, and in biological validation compared to gSVM-SCAD (**see supplementary material** for biological validation results). We have only focused on static gene expression data; however, our approach can be implemented with modification to time-course or survival gene expression data.

Acknowledgement:

We would like to thank Universiti Teknologi Malaysia for supporting this research by the UTM GUP research grant (Vot number: QJ130000.7107.01H29).

References:

[1] Wang X *et al. Curr Opin Biotechnol*. 2008 **19**: 482 [PMID: 18725293]
[2] Chen X *et al. Bioinformatics* 2008 **24**: 2474 [PMID: 18753155]
[3] Tai F & Pan W. *Bioinformatics* 2007 **23**: 1775 [PMID: 17483507]

[4] Misman MF *et al. ICIC Express Letters, Part B: Applications* 2010 1: 131.
[5] Fan J & Li R. *Journal of the American Statistical Association* 2001 **96**: 456
[6] Shi W *et al. Stat interface*. 2008 **1**: 137 [PMID: 18852828]
[7] Battacharjee A *et al. Proc Natl Acad Sci U S A*. 2001 **98**: 13790 [PMID: 11707567]
[8] Enerson BE *et al. Toxicol Pathol*. 2006 **34**: 27 [PMID: 16507541]
[9] Schwarz G. *Ann Statist*. 1978 **6**: 461
[10] Wang H *et al. Biometrika* 2007 **94**: 553 [PMID: 19343105]
[11] Wang H *et al. Journal of the Royal Statistical Society. Series B: Statistical Methodology* 2009 **71**: 3
[12] Wahba G *et al. Advances in Large Margin Classifiers, MIT Press* 2000 297-309.
[13] Mohamad MS *et al. IEEE Tran INF Technol Biomed*. 2011 [PMID: 21914573]
[14] Becker N *et al. Bioinformatics* 2009 **25**: 1711 [PMID: 19398451]
[15] Pang H *et al. Bioinformatics* 2006 **22**: 2028 [PMID: 16809386]

Edited by P Kangueane
Citation: Misman *et al.* Bioinformation 7(4): 169-175 (2011)

BIOINFORMATION

open access

Supplementary material:

The formula of GACV is shown below:

$$GACV(\lambda) = \frac{1}{n}\sum_{n=1}^{n}(1 - y_i\beta_{\lambda i}) + DF_\lambda$$

Where n, β, and λ refers to number of samples, coefficients of the hyperplane, and tuning parameter, respectively while DF_λ is the degree of freedom of parameter λ, where

$$DF_\lambda = \frac{1}{n}\left[2\sum_{y_i f(x_i)<-1}\frac{\alpha_{\lambda i}}{2n\lambda}\cdot\|K(\cdot,x_i)\|^2_{Hk} + \sum_{y_i f(x_i)\in[-1,1]}\frac{\alpha_{\lambda i}}{2n\lambda}\cdot\|K(\cdot,x_i)\|^2_{Hk}\right]$$

Where $\frac{\alpha_{\lambda i}}{2n\lambda} = \frac{f(x_{i+1})[y_{i+1}]-f(x_i)[y_i]}{y_{i+1}-y_i}$ and $\|K(\cdot,x_i)\|^2_{Hk}$ is the reproducing kernel hilbert space (RKHS) with SVM reproducing kernel K **[12].** Acording to Shi *et al.* **[6],** in the transformation akaike information criterion (AIC) to BIC, BIC replaces degree of freedom, DF_λ with $\left(\frac{\log n}{2}\right)DF_\lambda$. Therefore, the BGACV can be defined as,

$$BGACV(\lambda) = \frac{1}{n}\sum_{n=1}^{n}(1 - y_i\beta_{\lambda i}) + \frac{\log n}{2}DF_\lambda$$

The nearly optimal tuning parameter λ is obtained by minimizing the error rate from the BGACV.

Table 1: Gene expression datasets

Name	No. of samples	No. of genes	Class	Reference
Lung	86	7129	2(normal and tumor)	**[7]**
Canine	31	12473	2(control and drug induced dogs)	**[8]**

Table 2: A comparison of averages 10-fold CV accuracies from the top ten pathways with other supervised machine learning methods for the lung cancer and canine datasets.

Method	10-fold CV (%)	
	Canine	Lung cancer
gSVM-SCAD$_{BGACV}$	85.96	76.55
gSVM-SCAD (Misman *et al.*, 2010)	81.08	73.77
PathwayRF (Pang *et al., 2006*)	85.92	71.00
Neural Networks	82.33	70.39
KNN1	81.00	61.73
KNN3	81.00	64.80
LDA	76.62	63.24
L_1 SVM	72.33	55.14
SVM-SCAD	73.67	53.5

Note: The texts in **Bold** are the highest 10-fold CV accuracy. The texts in *italic* are the methods from the self-running experiment.

Biological Validation:

The five most significant pathways in the lung cancer dataset are (1) WNT signaling pathway, (2) AKAP95 role in mitosis and chromosome dynamics pathway, (3) Induction of apoptosis pathway, (4) Antisense pathway, and (5) Cycling of Ran in nucleocytoplasmic transport pathway as shown in **Table 3**. For the WNT signaling pathway, Mazieres *et al.* **[1]** reported that the pathway plays a significant role in the development of lung and other colorectal cancers. With respect to the second pathway, it also contributes to the development of lung cancer, since the AKAP95 protein plays an important role in cell mitosis **[2]**. The gSVM-SCAD identifies that induction of apoptosis pathway as one of the lung cancer related pathway, where this pathway has been reported by Lee *et al.* **[3]** as one of the contributor to the lung cancer development. For the Keratinocyte Differentiation pathway, Massion *et al.* **[4]** had shown that this pathway has contributed to the development of lung cancer. While for the Activation of Csk pathway, Masaki *et al.* **[5]** have reported that the activation of this pathway plays an important role in the development of lung cancer, with three genes marked as lung cancer genes. The selected genes for each pathway are also presented in **Table 3**.

Table 3: Selected genes from the top five pathways in the lung cancer dataset

Pathways	No. of genes	Selected gene(s)
WNT Signaling Pathway	24	**APC [1]**, **HNF1A [6]**, **CREBBP [7]**, **HDAC1 [8]**, **MYC [1]**, **WNT1 [9]**, **CSNK1A1** [10], **AXIN1** [1], **CSNK2A1 [10]**, **CTNNB1 [1]**, **GSK3B [1]**, **TLE1** [11], *PPARD*, *PPP2CA, TAB1*, *DVL1*
AKAP95 role in mitosis and chromosome dynamics	10	**DDX5 [12]**, **PRKACB [10]**, **CDK1 [13]**, **CCNB1 [14]**, *PPP2CA, PRKAR2B, PRKAR2A*
Induction of apoptosis	36	**FADD [15]**, **TNFSF10 [16]**, **CASP7 [3]**, **BCL2 [17]**, **BIRC3 [18]**, **CASP9 [3]**, **TRAF [19]**, **BIRC [20]**, **CASP8 [3]**, **CASP3 [3]**, **TNFRSF25 [21]**, **CASP10 [3]**, **RARA [22]**, **CASP6 [3]**, *TRADD*, *RELA*, *DFFA, RIPK1*
Keratinocyte Differentiation	45	**Hras [23]**, **FASLG [24]**, **ETS1 [25]**, **JUN [26]**, **MAPK14 [27]**, **BCL2 [28]**, **MAPK8 [29]**, **EGFR [30]**, **RAF1 [31]**, **FAS [32]**, *RELA, PRKCQ, PPP2CA, MAP2K6, NFKB1,* CEBPA
Activation of Csk	30	**PRKACB [10]**, **CREBBP [7]**, **HLA-DQB1 [10]**, *CD247, IL23A, PRKAR1B*, *GNGT1*, *CD3D*, *CD3E*

Note: The genes in *italic* text are uninformative. The genes in **bold** text are genes that directly related to the lung cancer.

For the canine dataset as shown in **Table 4**, the gSVM-CAD$_{BGACV}$ has been detected. Four pathways are similar to the pathways that are detected by Pang *et al.* **[33];** the pathways are Aminoacyl-tRNA biosynthesis pathway, Granzyme A mediated Apoptosis Pathway, TGF-beta signaling pathway, and Antisense Pathway. Hence, all these four pathways are considered significant based on the literature. For the Starch and Metabolism pathway, it may not seem directly related to vascular pathology, but it contained two genes that related to the metastasis for gene RUVBL2 **[34]** and inflammation for UGDH **[35].** This makes it possible that this pathway may relate to the vascular pathology and thus prompting biologists to conduct further research on this pathway. The selected genes for each pathway are presented in **Table 4**. Fifteen genes were selected, with thirteen genes detected as informative.

Table 4: Selected genes from the top five pathways in the canine dataset

Pathways	No. of genes	Selected gene(s)
TGF-beta signaling pathway	40	**RHOA [33]**, **BMP4 [33]**, **SP1 [33]**
Starch and sucrose metabolism	59	*TESK1*, **RUVBL2 [34]**, **UGDH [35]**
Aminoacyl-tRNA biosynthesis	19	**NARS [33]**, **FARS [33]**, **FARSA [33]**
Granzyme A mediated Apoptosis Pathway	5	*SET*, **APEX1 [33]**, **DFFB [33]**
Antisense Pathway	34	**SFPQ [33]**, **MATR3 [33]**, **ADAR [33]**

Note: The genes in *italic* text are uninformative. The genes in **bold** vascular injury in coronary arteries related.

References:

[1] Mazieres J *et al. Cancer Lett.* 2005 **222**: 1 [PMID: 15837535]
[2] Collas P *et al. J Cell Biol.* 1999 **147**: 1167 [PMID: 10601332]
[3] Lee SY *et al. J Thorac Oncol.* 2010 **5**: 1152 [PMID: 20661084]
[4] Massion PP *et al. Cancer Res.* 2003 **63**: 7113 [PMID: 14612504]
[5] Masaki T *et al. Eur J Cancer.* 2003 **39**: 1447 [PMID: 12826049]
[6] Lanzafame S *et al. Pathologica.* 2006 **98**: 640 [PMID: 17285841]
[7] Tillinghast GW *et al. Genes Chromosomes Cancer.* 2003 **37**: 121 [PMID: 12696060]
[8] Sasaki H *et al. Lung Cancer.* 2004 **46**: 171 [PMID: 15474665]
[9] Huang CL *et al. Eur J Cancer.* 2008 **44**: 2680 [PMID: 18790633]
[10] Battacharjee A *et al. Proc Natl Acad Sci U S A.* 2001 **98**: 13790 [PMID: 11707567]
[11] Jagdis A *et al. Am J Surg Pathol.* 2009 **33**: 1743 [PMID: 19809272]
[12] Su LJ *et al. BMC Genomics.* 2007 **8**: 140 [PMID: 17540040]
[13] Hsu TS *et al.Cancer Chemother Pharmacol.* 2008 **62**: 799 [PMID: 18193228]
[14] Kosacka M *et al. Pol Merkur Lekarski.* 2010 **28**: 117 [PMID: 20369739]
[15] Bhojani MS *et al. Cell Cycle.* 2005 **4**: 1478 [PMID: 16258269]
[16] Sun W *et al. Cancer Lett.* 2004 **212**: 83 [PMID: 15246564]
[17] Nhung NV *et al. Cesk Patol.* 1999 **35**: 117 [PMID: 10677910]
[18] Ekedahl J *et al. Exp Cell Res.* 2002 **279**: 277 [PMID: 12243753]
[19] Li X *et al. Nature* 2002 **416**: 345 [PMID: 11907583]
[20] Kang HG *et al. Cancer Genet Cytogenet.* 2008 **180**: 6 [PMID: 18068526]
[21] Anglim PP *et al. Mol Cancer.* 2008 **7**: 62 [PMID: 18616821]
[22] Wan H *et al. Cancer Res.* 2001 **61**: 556 [PMID: 11212249]
[23] Araki T *et al. J Mol Diagn.* 2010 **12**: 118 [PMID: 20007840]
[24] Myong NH. *J Korean Med Sci.* 2005 **20**: 770 [PMID: 16224150]
[25] Sasaki H *et al. J Surg Res.* 2001 **101**: 242 [PMID: 11735282]
[26] Gamou S & Shimizu N. *Cell Struct Funct.* 1994 **19**: 195 [PMID: 7529661]

[27] Chien PS *et al. Biochem Biophys Res Commun.* 2006 **339**: 562 [PMID: 16300728]
[28] Tas F *et al. Med Oncol.* 2005 **22**: 139 [PMID: 15965276]
[29] Chuang SM *et al. Carcinogenesis* 2003 **24**: 7 [PMID: 12538343]
[30] Sasaki H *et al. Int J Clin Oncol.* 2003 **8**: 79 [PMID: 12720099]
[31] Cekanova M *et al. Cancer* 2007 **109**: 1164 [PMID: 17315157]
[32] Visca P *et al. Anticancer Res.* 2004 **24**: 4169 [PMID: 15736468]
[33] Pang H *et al. Bioinformatics* 2006 **22**: 16
[34] Diop SB *et al. EMBO Rep.* 2008 **9**: 260 [PMID: 18259215]
[35] Uzuki M *et al. Anal Quant Cytol Histol.* 1999 **21**: 75 [PMID: 10068780]

