

Systems biology

Combining tissue transcriptomics and urine metabolomics for breast cancer biomarker identification

Hojung Nam¹, Bong Chul Chung², Younghoon Kim¹, KiYoung Lee³ and Doheon Lee^{1,*}¹Department of Bio and Brain Engineering, KAIST, 373-1 Guseong-dong, Yuseong-gu, Daejeon 305-701,²Bioanalysis and Biotransformation Research Center, KIST, Chengryang, Seoul 130-605 and ³Department of Biomedical Informatics, Ajou University School of Medicine, Suwon 443-749, Korea

Received on May 27, 2009; revised on September 8, 2009; accepted on September 21, 2009

Advance Access publication September 25, 2009

Associate Editor: Limsoon Wong

ABSTRACT

Motivation: For the early detection of cancer, highly sensitive and specific biomarkers are needed. Particularly, biomarkers in bio-fluids are relatively more useful because those can be used for non-biopsy tests. Although the altered metabolic activities of cancer cells have been observed in many studies, little is known about metabolic biomarkers for cancer screening. In this study, a systematic method is proposed for identifying metabolic biomarkers in urine samples by selecting candidate biomarkers from altered genome-wide gene expression signatures of cancer cells. Biomarkers identified by the present study have increased coherence and robustness because the significances of biomarkers are validated in both gene expression profiles and metabolic profiles.

Results: The proposed method was applied to the gene expression profiles and urine samples of 50 breast cancer patients and 50 normal persons. Nine altered metabolic pathways were identified from the breast cancer gene expression signatures. Among these altered metabolic pathways, four metabolic biomarkers (Homovanillate, 4-hydroxyphenylacetate, 5-hydroxyindoleacetate and urea) were identified to be different in cancer and normal subjects ($p < 0.05$). In the case of the predictive performance, the identified biomarkers achieved area under the ROC curve values of 0.75, 0.79 and 0.79, according to a linear discriminate analysis, a random forest classifier and on a support vector machine, respectively. Finally, biomarkers which showed consistent significance in pathways' gene expression as well as urine samples were identified.

Contact: dhlee@biosoft.kaist.ac.kr

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 INTRODUCTION

The altered metabolic activities of cancer cells have been observed in many studies. It has been reported that most cancer cells predominantly produce energy by glycolysis in the cytosol (Gatenby and Gillies, 2004; Kim and Dang, 2006; Moreno-Sanchez *et al.*, 2007). In addition, it has been shown that cancer cells demonstrate an increased rate of protein turnover as well as an increase in lipolysis, which is the breakdown of fat stored in fat cells

(Al-Majid and Waters, 2008; Argiles *et al.*, 2007; Yeh *et al.*, 2006). It has also been reported that human colon cells show deregulation of the TCA cycle and amino acid turnover (Denkert *et al.*, 2008). Based on the perception of cancer metabolism, groups of scientists have devoted much effort to the identification of metabolic biomarkers for non-invasive cancer screening. Recently, many researchers have identified sensitive cancer biomarkers with the technical development in mass spectrometry-based metabolomic studies. Deng *et al.* (2004) reported hexanal and heptanal in serum as biomarkers for lung cancer. Philips *et al.* (2003, 2006) identified decanes and heptanes for breast cancer in breath samples with a sensitivity of 88.2% and a specificity of 73.8%. Henneages *et al.* (2009) reported 44 pair-wise ratios of RNA metabolites for breast cancer urinary tests. In addition, Sreekumar *et al.* (2009) identified highly increased levels of sarcosine during prostate cancer progression to metastasis.

Mass spectrometry-based metabolomic biomarker studies can be categorized into two classes, targeted and untargeted approaches. In targeted approaches, candidate biomarkers which appear to be related to cancer specific functions are selected in advance based on prior knowledge. Once the candidate biomarkers are selected, candidate biomarkers in mass spectrum data are identified and the discriminative powers of the candidate biomarkers are tested. Additionally, in non-targeted approaches, statistical or multivariate analysis is applied to the aligned mass spectrum data, and biomarkers which show significant differences between cancer and normal patients are then identified. However, there remain many difficulties in the identification of biomarkers in both approaches. In targeted approaches, if the candidate biomarkers were initially selected incorrectly, biomarker identification may easily fail, even if there are potential biomarkers with strong discriminative powers in the samples (Beckmann *et al.*, 2008). In non-targeted approaches, it is difficult to identify consistent biomarkers, as metabolic phenotypes are easily affected by various environmental and lifestyle factors such as gender, age and diet (Holmes *et al.*, 2008; Li *et al.*, 2008). In addition, from a statistical point of view, it is difficult to identify robust biomarkers due to the inherent low sample-to-variable ratio of mass spectrum data.

It is expected that the development of microarray technologies and large-scale metabolic pathway structure information can assist with candidate biomarker selection in targeted approaches. Recently, gene-set analysis methods for evaluating the activities of biological

*To whom correspondence should be addressed.

pathways were introduced and applied to metabolic pathways (Dinu *et al.*, 2007; Subramanian *et al.*, 2005). Gene-set enrichment analysis (GSEA) is one of the most well-established gene-set approaches that determines whether an a priori defined set of genes concordantly shows statistically significant differences between two phenotypes (Subramanian *et al.*, 2005). However, because there is typically at least one rate-determining step that determines the overall activity of a metabolic pathway, not all metabolic genes of a pathway consistently show significant transcriptional changes even if the activities of pathway are altered (Hall *et al.*, 2007; Nakamura *et al.*, 2005). To overcome this limitation, the authors recently proposed a new scoring method for measuring the altered activities of metabolic pathways and showed its effectiveness (Nam *et al.*, 2009).

The present study proposes a systematic method for the identification of metabolic biomarkers for urinary tests that selects candidate biomarkers from altered genome-wide gene expression signatures. Our basic idea is combining gene expression profiles of cancer cells and metabolomic profiles of urine to identify coherent urinary biomarkers that originate from cancer cells and are then excreted in urine. First, in order to identify altered breast cancer metabolism, we use a computational method that identifies genome-wide altered metabolic pathways by analyzing gene expression profiles and predefined functional units of metabolic pathways in KEGG maps (Nam *et al.*, 2009). This scoring method observes changed expression levels of metabolic genes that show significant expression changes in breast cancer compared to normal subjects. Second, with the identification of the metabolic pathways to be altered, candidate biomarkers, which are substrates of the altered metabolism, are selected. Once the candidate biomarkers are selected, the discriminative powers of the selected candidate biomarkers are tested with the collected gas chromatography-mass spectrometry (GC-MS) data of urine samples. Finally, a set of biomarkers for urinary cancer screening tests is given.

2 MATERIALS AND METHODS

Figure 1 provides an overview of the proposed method. The significance of a change in the level of genes is extracted by applying *t*-tests to microarray data. The transcription factor information of enzymes is merged with the information of the metabolic pathways. With these two sources of information, the alteration scores of the pathways are measured. Once the altered metabolic pathways are identified, the discriminative powers of candidate biomarkers are evaluated and sets of biomarkers for breast cancer urinary tests are identified.

2.1 Data sets for identification of altered metabolism

In this work, three different sources of data were used to identify altered metabolism (see Fig. 1). First, gene expression data sets of breast cancer and normal subjects obtained from the NCBI GEO database were used (Table 1). To avoid obtaining biased signatures of a certain histological type of cancer, we used gene expression data sets consisted of more than one histological type. Second, the transcriptional regulation information of *Homo Sapiens* from the TRANSFAC® database (Matys *et al.*, 2003, 2006) was used. A total of 364 instances of protein-DNA binding information were obtained that had explicit regulation annotation terms (e.g. activator, inhibitor). Last, 112 instances of metabolic pathway information from the KEGG Pathway database were used (Kanehisa and Goto, 2000; Kanehisa *et al.*, 2006, 2008).

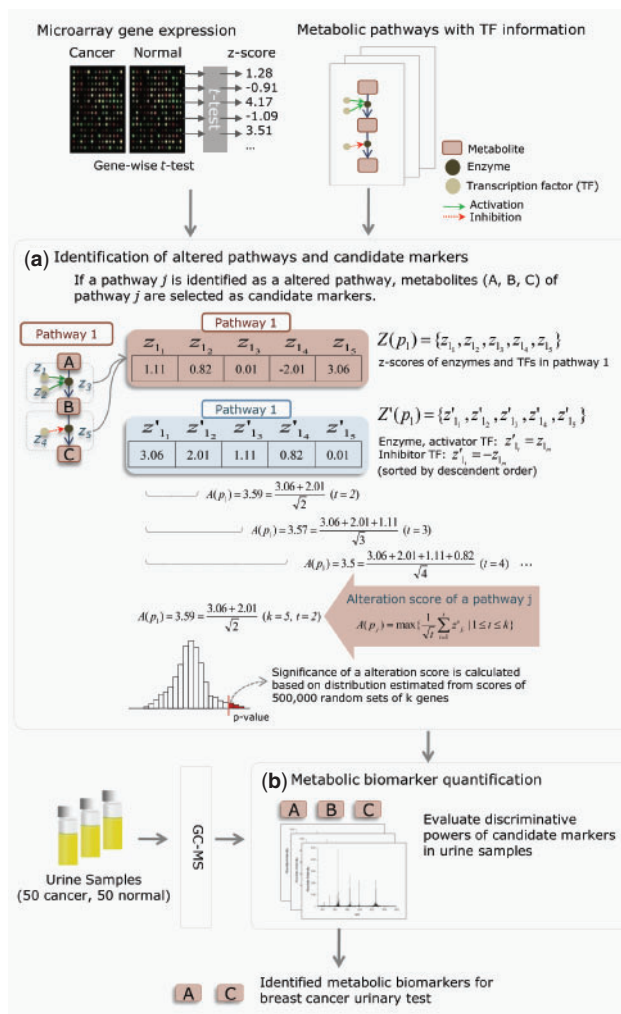


Fig. 1. Method overview. (a) Identification of altered pathways and candidate biomarkers. Using microarray gene expression profiles, the standardized *z*-scores of genes are obtained by *t*-tests. With the obtained *z*-scores and the metabolic pathway information, the alteration scores of each metabolic pathway were measured. An alteration score of a pathway is an aggregated form of the expression scores of genes in a pathway that maximizes the aggregated score. (b) Metabolic biomarker identification. By evaluating the discriminative power and classification performance of candidate biomarkers, sets of urinary biomarkers for cancer screening are reported.

Table 1. Used breast cancer gene expression data sets

Data set (GEO ID)	Number of samples (cancer/normal)	Histological types
GSE3744	40/7	Basal; non-basal; BRCA1-associated
GSE10810	31/27	Invasive ductal carcinoma; invasive lobular carcinoma

2.2 Urine sample collection

Urine samples were collected from female breast cancer patients at various cancer stages with a mixture of histological types of invasive ductal carcinoma (IDC), invasive lobular carcinoma (ILC), ductal carcinoma in situ (DCIS) ($n = 50$, age 47.6 ± 7.89 years). Samples of healthy subjects as normal controls ($n = 50$, age 46.64 ± 7.38 year) were collected at the Samsung and Hanyang University Medical Centers (Seoul, Korea). All study subjects underwent the same diagnostic procedures, i.e. a physical breast examination, a mammography, and ultrasonography as detailed by the American Joint Committee on Cancer Staging. Additionally, none of the subjects had undergone chemotherapy. The normal sex- and age-matched controls had no evidence of benign or malignant breast disease. Detailed clinical information pertaining to the 50 cancer patients are listed in Supplementary Table 1 in Supplementary Appendix 1.

2.3 Sample preparation for GC-MS analysis

Urinary metabolites were prepared under four different conditions. In the first condition, a urine sample (1 ml) was loaded into Strata-X cartridge (60 mg, 3 ml; Phenomenex, Torrance, CA), washed with distilled water (1 ml) and extracted with 4 ml of methanol. The eluant was divided into halves. One half of the eluant (2 ml) was evaporated and dried in a desiccator over 30 min. In the second condition, the other half was evaporated and dissolved with 1 ml of 0.2 M acetate buffer (pH 5.2) and hydrolyzed with β -glucuronidase/arylsulfatase (50 μ l) from Helix Pomatia (Roche, Mannheim, Germany) at 55°C for 3 h. After cooling, urinary metabolites were extracted with 5 ml of diethyl ether by mechanical shaking for 20 min and centrifugation at 2500 rpm for 5 min. The separated organic layer was evaporated under nitrogen and kept in the desiccator for 30 min. In the third condition, the remaining aqueous layer had 200 μ l of 3 M HCl added to it to adjust it to a pH range of 1–2. It was then extracted with 5 ml of diethyl ether. The separated organic extract was evaporated and dried. In the fourth condition, the remaining aqueous layer had 0.73 g of K_2CO_3 added to it to adjust it to a pH range of 10–11. It was then extracted with 5 ml of diethyl ether and dried. All dried extracts were derivatized by 50 μ l of MSTFA/TMSI/TMCS (100:2:5, v/v/v) mixture at 60°C for 15 min and injected into a GC-MS system.

2.4 Instrumental conditions

All samples prepared were separated through a Ultra-1 capillary column (25 m \times 0.2 mm ID, 0.33 μ m film thickness; Agilent, Palo Alto, CA) and analyzed by Thermo Finnigan GC-MS systems that consisted of a model Trace 2000 GC and Polaris Q mass-selective detector in a scan range of m/z 50–800 (Thermo Finnigan, Waltham, MA).

2.5 Quantification of target metabolites

The intensities of the metabolites in the urine samples were quantified via AMDIS 2.6 (Automated Mass Spectral Deconvolution and Identification System) (Stein, 1999) and reference libraries (NIST 02 and Wiley 7 MS libraries). Once the intensities of candidate metabolic markers were quantified, the intensities of a candidate marker were aligned across all samples (both 50 cancer and 50 normal samples). Finally candidate markers that were missing no more than 25% of their values were selected. After the alignment, the missing values of the intensities were input using the *knn* imputation function, which is implemented in the *R* ‘*impute*’ package.

2.6 Identifying altered metabolic pathways

To calculate the alteration levels of a metabolic pathway, the significance of the differential expression of each gene under cancer was initially measured. To do this, the Student's *t*-test was used to evaluate the significance of the change of a gene. The obtained *P*-value (p_i) of each gene ($gene_i$) was converted to a *z*-score (z_i) using the inverse normal cumulative distribution function (θ^{-1}). A *z*-score follows a standard normal distribution, and a higher

z-score indicates more significantly induced or reduced gene expression in cancer.

$$z_i = \theta^{-1}(1 - p_i)$$

A pre-published scoring method by the authors was then applied to measure the altered activities of the metabolic pathways to the breast cancer gene expression data set. Details of this method for identifying altered metabolic pathways are explained in the Supplementary Material.

2.7 Classification evaluation

To evaluate the predictive performance of the identified biomarkers, supervised machine learning classification techniques were used. These techniques involved the use of three classification algorithms implemented in *R* packages: a linear discriminative analysis (LDA), the random forest classifier (RF) and the support vector machine (SVM). The data set was first randomly split onto training and testing sets. Five-fold cross validation was performed on the training set and the generalization was obtained on an unseen testing set. Five-fold cross validation was repeated 1000 times to generalize the performance of the classifier. The final classification performance is reported as the averaged values of the area under the ROC curve (AUC) on the test set. To test the performance of multiple biomarkers, the biomarkers were initially ranked according to their AUC values, and every classifier was then built by adding a biomarker serially in decreasing order of the AUC value.

3 RESULTS

3.1 Identified altered metabolism and hallmarks of cancer metabolism

The proposed method was applied to two breast cancer gene expression profiles (Table 1). From each data set, 14 and 25 metabolic pathways were identified to be altered in GSE3744, GSE10810, respectively (Fig. 2a). To avoid obtaining biased signatures of a certain histological type of cancer, pathways which are commonly identified in both expression data sets are selected for further analysis (Fig. 2b). The genes and their corresponding reactions which maximized the alteration score in each pathway are fully listed in Supplementary Table 1 in Supplementary Appendix 2. Table 2 lists the altered metabolic pathways in the results. Nine altered metabolic pathways were commonly identified from the two breast cancer gene expression data sets. It should be noted that some correspondence exists between the present findings and previously known cancer hallmarks. It has been reported that the activity of the pyrimidine biosynthetic pathway in MCF7 breast cancer cells was 4.4-fold higher than that in normal MCF10A breast cells (Sigoillot *et al.*, 2004). In the case of tyrosine metabolism, it was reported that the signal transducer and activator of transcription 3 (Stat3) is constitutively tyrosine-phosphorylated in ~50% of primary breast carcinomas (Berishaj *et al.*, 2007). It has also been shown that the plasma levels of total tryptophan were significantly lower in patients with breast cancer than in women with benign breast cancer (Poulter *et al.*, 1985).

3.2 Comparative analysis of candidate metabolic biomarkers from altered metabolism

Based on the nine identified metabolic pathways to be altered, candidate biomarkers which are substrates of the altered pathways were selected. When selecting the candidates, the metabolites having the standard mass peaks of derivatives in a reference library (NIST 02 and Wiley 7 MS libraries) were chosen for the

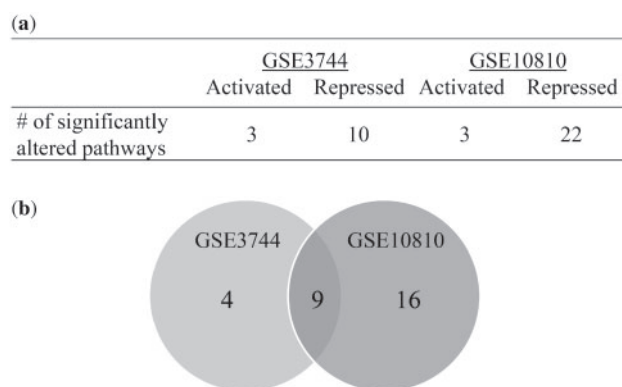


Fig. 2. Results from two gene expression data sets. (a) The number of significantly altered metabolic pathways identified by the proposed method ($FDR \leq 5\%$). (b) Venn diagram of overlap between identified altered pathways. The figure shows the overlap between the altered pathways for both data sets.

Table 2. Identified metabolic pathways as altered in two breast cancer gene expression data sets ($FDR \leq 0.05$)

Name	Alteration type	Number of genes
Pyrimidine metabolism	Activated	90
Purine metabolism	Activated	146
Valine, leucine and isoleucine degradation	Repressed	46
Tyrosine metabolism	Repressed	48
Arachidonic acid metabolism	Repressed	59
Metabolism of xenobiotics by cytochrome P450	Repressed	60
Butanoate metabolism	Repressed	38
Tryptophan metabolism	Repressed	50
Fatty acid metabolism	Repressed	44

peak matching tests with the urine samples. The total number of candidate metabolic biomarkers in breast cancer is 128 (see Supplementary Table 2 in Supplementary Appendix 1). Once the candidate biomarkers were selected, the mass peaks of the candidate biomarkers in the GC-MS data were identified with AMDIS 2.6 software. During the peak detection, metabolites that did not match the quality criteria underlying the peak detection algorithm at a given sample are resulted in missing quantified values. In the present study, we took into account metabolites those were consistently detected in at least 75% of samples in order to maximize the quality of urine profiles. Finally, the intensities of 15 metabolites for breast cancer were quantified across 100 urine samples (see Supplementary Table 3 in Supplementary Appendix 1). The sums of the peak areas of the urine samples of 50 breast cancer patients and 50 normal samples were analyzed. The differences in the intensities of the candidate biomarkers were then compared between cancer and normal groups in a two-tailed *t*-test, and six metabolic biomarkers (hippurate, homovanillate, benzoate, 4-hydroxyphenylacetate, 5-hydroxyindoleacetate and urea) were finally determined to be different ($P < 0.05$) (see Fig. 3, and

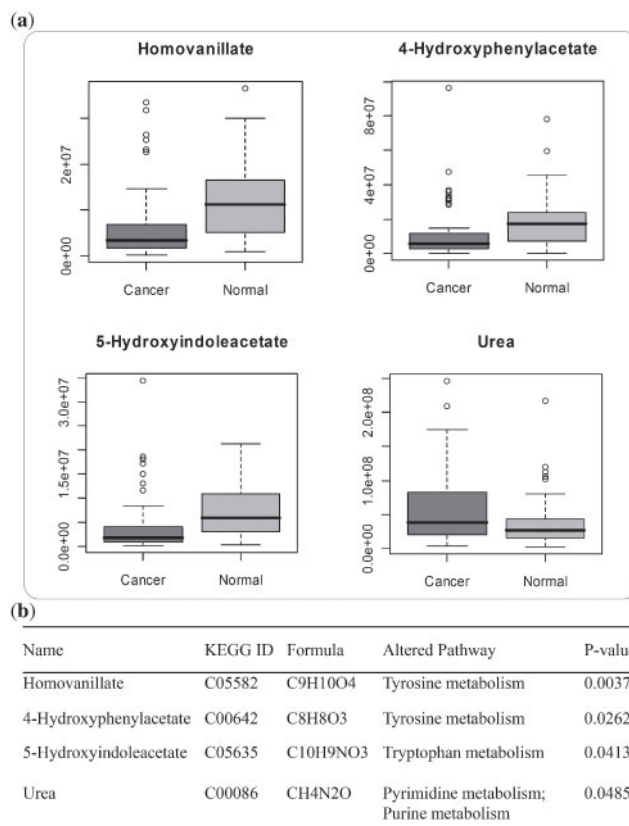


Fig. 3. Four selected breast cancer biomarkers in urine samples. (a) Box-whisker plots intensities of selected biomarkers in the urine samples ($P < 0.05$). (b) The detailed information of the selected biomarkers.

Supplementary Figs 1–4 in Supplementary Appendix 2). In next section, the predictive performance of selected four biomarkers is evaluated.

3.3 Classification of cancer versus normal with selected biomarkers

The urine samples consisted of 50 cancer samples and 50 normal samples, and the data was randomly split into 40 samples of training and 10 samples of testing sets for each phenotype. For an 80:20 train-test set, 5-fold cross validation was performed on the training set, and the generalization was obtained on an unseen testing set. The final classification performance is reported as the averaged AUC values on the test set for each classifier.

Figure 4 shows the classification performance of the selected biomarkers with the three classifiers. The performance of each selected biomarker was tested first, and the combinatorial effect of multiple biomarkers was then tested. To test the performance of multiple biomarkers, the biomarkers were initially ranked according to their AUC values. Each classifier was then formulated by adding a biomarker serially in decreasing order of its AUC value. As the figure shows, the multiple biomarkers achieved better classification results compared to single biomarkers across the three classifiers (Fig. 4a–c). For example, multiple biomarkers achieved a highest AUC value of 0.75 with LDA, whereas a single biomarker achieved the highest AUC value of 0.74. Sets of multiple biomarkers which achieved the

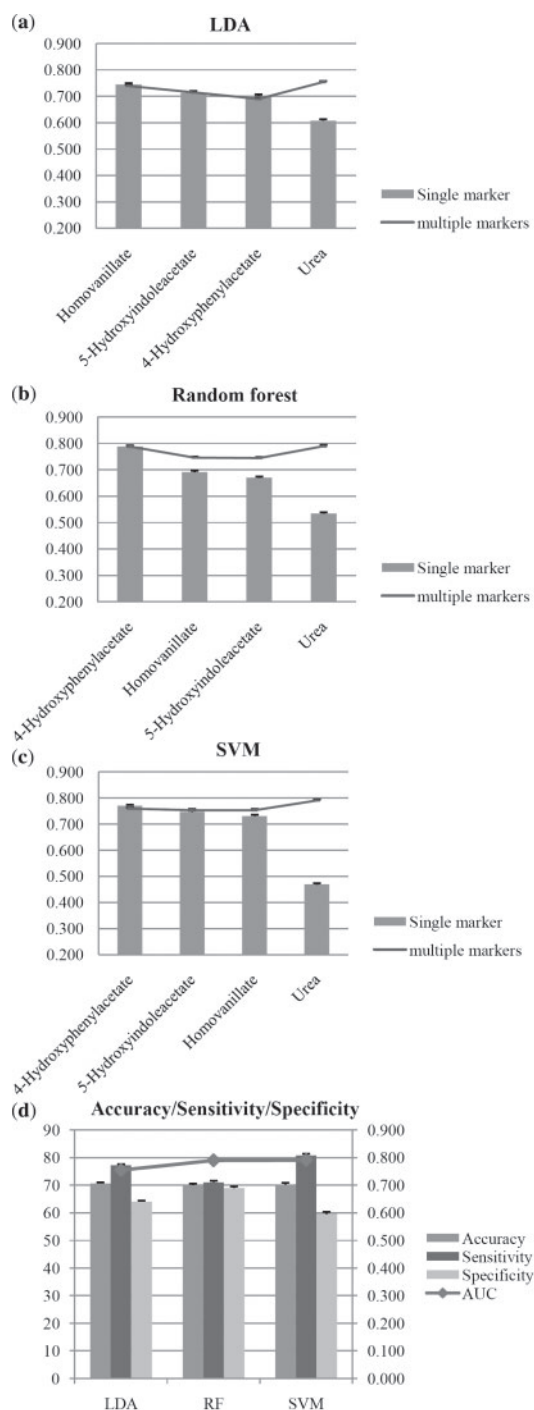


Fig. 4. The classification performance of LDA, RF and SVM with the four selected biomarkers: (a–c) The bar in each plot indicates the mean and the standard error of the mean (SEM) of the AUC value for each single biomarker. Biomarkers are ranked in decreasing order of their AUC values on the x-axis. The line in each plot indicates the mean and the SEM of the AUC values of accumulated multiple biomarkers. (d) Accuracy, sensitivity, specificity rates of multiple biomarkers at the highest AUC value [accuracy = $(TP + TN)/(TP + FP + FN + TN)$, sensitivity = $TP/(TP + FN)$, specificity = $TN/(FP + TN)$].

best performance showed AUC values of 0.75, 0.79 and 0.79 with LDA, Random forest, SVM, respectively. Accuracy, sensitivity and specificity rates of multiple biomarkers were measured when sets of multiple biomarkers achieve the highest AUC value (Fig. 4d). Sensitivity is a rate of cancers that are correctly identified as cancer. Specificity is a rate of normal that are correctly identified as normal. As the figure shows, identified biomarkers showed better sensitivity rates than specificity rates across all three classifiers. As screening test is intended to have a high sensitivity, it is important to compare the sensitivity performance of the identified urine markers and mammography which is one of the standards for breast cancer screening tests. According to previous researches, the sensitivity of mammography is reported as around 70–80% (Houssami *et al.*, 2002; Jensen *et al.*, 2006). The identified biomarkers showed similar sensitivity performance compared to mammography. However the urinary test with the biomarkers could be more beneficial because mammography has potential risk of radiation exposure. This result shows that the identified metabolic biomarkers that originate from gene expression signatures have discriminative power for a urinary test in the screening of cancer patients.

4 DISCUSSION

4.1 Correlation between candidate biomarkers and cancer progression

To determine whether the candidate markers can be indices of cancer progression, the correlations between the intensities of a candidate marker and the cancer stages were evaluated. First, the averages of the intensities of a candidate marker in each cancer stages were calculated. Subsequently, the Pearson's correlation coefficient of a candidate marker was estimated against the progression of the cancer (see Supplementary Table 4 in Supplementary Appendix 1). The results show that among 15 candidate biomarkers, thymine levels showed a correlation with cancer progression (Pearson's correlation coefficient: 0.852). The thymine levels across cancer progression are depicted in Figure 5.

4.2 The origin of identified biomarkers

To ensure that the identified urinary metabolic biomarkers originated from breast cancer cells, it is necessary to confirm whether the biomarkers could be secreted to blood from cancer cells and then excreted in urine. Therefore, the biofluid locations of the identified biomarkers were checked by referring to the Human Metabolome Database (HMDB) (Wishart *et al.*, 2007). Table 3 shows a detailed description of the identified biomarkers and their biofluid location information. As the tables shows, all identified biomarkers except benzoate are identified in both blood and urine biofluid. Therefore, this result is evidence that the altered metabolic signatures of cancer cells can assist with the identification of urinary biomarkers.

5 CONCLUSION

The present study proposes a systematic method for the identification of metabolic biomarkers for urinary tests that selects candidate biomarkers from altered genome-wide gene expression signatures. Altered metabolic signatures of breast cancer were identified through statistical tests using gene expression and KEGG metabolic pathways. Once the altered metabolic pathways are selected, the

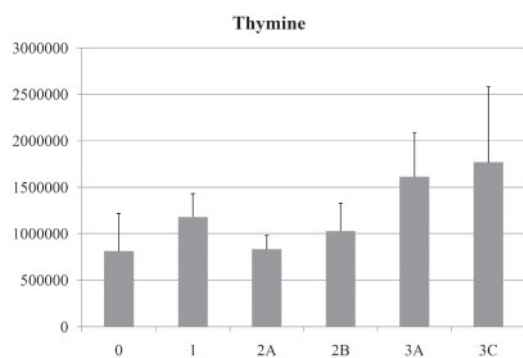


Fig. 5. The intensities of thymine across cancer progression: The *x*-axis indicates cancer stages. The bar denotes the average and the SEM of intensities.

Table 3. Four identified biomarkers and their biofluid location information

Marker	Biofluid location
Homovanillate	Blood, cerebrospinal fluid, urine
4-Hydroxyphenylacetate	Blood, cerebrospinal fluid, urine; saliva
5-Hydroxyindoleacetate	Blood, cerebrospinal fluid, urine
Urea	Blood, cellular cytoplasm, cerebrospinal fluid, Urine

discriminative powers of selected candidate biomarkers are tested with the collected gas chromatography-mass spectrometry (GC-MS) data of urine samples. The proposed method was applied to the gene expression profiles and urine samples of breast cancer patients and normal persons. Nine altered metabolic pathways were identified from the breast cancer gene expression signatures. Finally, four metabolic biomarkers (Homovanillate, 4-hydroxyphenylacetate, 5-hydroxyindoleacetate and urea) were identified to be different in cancer and normal subjects ($P < 0.05$). In the present study, we proposed a systematic method for the identification of metabolic biomarkers for urinary tests that selects candidate biomarkers from altered genome-wide gene expression signatures. To avoid obtaining biased signatures of a certain histological type of cancer, we used gene expression data sets consisted of various histological type. Also, in order to find out biomarkers that are not bias to specific patients, we used the urine and gene expression data from different cohort sets. We expect that the identified biomarkers by our method are more coherent and robust compared to biomarkers identified by conventional mass spectrometry-based metabolomic studies because the biomarkers' significances are validated in two independent data sets.

ACKNOWLEDGEMENTS

The authors thank the CHUNG Moon Soul Center for BioInformation and BioElectronics for providing the research facilities.

Funding: Korean Systems Biology Program (No. M10309020000-03B5002-00000) and the National Research Lab. Program (ROA-2005-000-10094-0) from the Ministry of Education, Science and

Technology through the Korea Science and Engineering Foundation. Also, supported from the KISTI super-computing project (K-09-L03-C03-S02).

Conflict of Interest: none declared.

REFERENCES

- Al-Majid,S. and Waters,H. (2008) The biological mechanisms of cancer-related skeletal muscle wasting: the role of progressive resistance exercise. *Biol. Res. Nurs.*, **10**, 7–20.
- Argiles,J.M. et al. (2007) Mechanisms to explain wasting of muscle and fat in cancer cachexia. *Curr. Opin. Support. Palliative Care*, **1**, 293–298.
- Beckmann,M. et al. (2008) High-throughput, nontargeted metabolite fingerprinting using nominal mass flow injection electrospray mass spectrometry. *Nature Protocols*, **3**, 486–504.
- Berishaj,M. et al. (2007) Stat3 is tyrosine-phosphorylated through the interleukin-6/glycoprotein 130/Janus kinase pathway in breast cancer. *Breast Cancer Res.*, **9**, R32.
- Deng,C. et al. (2004) Development of headspace solid-phase microextraction with on-fiber derivatization for determination of hexanal and heptanal in human blood. *J. Chromatogr.*, **813**, 47–52.
- Denkert,C. et al. (2008) Metabolite profiling of human colon carcinoma—deregulation of TCA cycle and amino acid turnover. *Mol. Cancer*, **7**, 72.
- Dinu,I. et al. (2007) Improving gene set analysis of microarray data by SAM-GS. *BMC Bioinformatics*, **8**, 242.
- Gatenby,R.A. and Gillies,R.J. (2004) Why do cancers have high aerobic glycolysis? *Nat. Rev. Cancer*, **4**, 891–899.
- Hall,R.K. et al. (2007) Insulin represses phosphoenolpyruvate carboxykinase gene transcription by causing the rapid disruption of an active transcription complex: a potential epigenetic effect. *Mol. Endocrinol.*, **21**, 550–563.
- Henneges,C. et al. (2009) Prediction of breast cancer by profiling of urinary RNA metabolites using Support Vector Machine-based feature selection. *BMC Cancer*, **9**, 104.
- Holmes,E. et al. (2008) Human metabolic phenotype diversity and its association with diet and blood pressure. *Nature*, **453**, 396–400.
- Houssami,N. et al. (2002) The comparative sensitivity of mammography and ultrasound in women with breast symptoms: an age-specific analysis. *Breast*, **11**, 125–130.
- Jensen,A. et al. (2006) Performance of clinical mammography: a nationwide study from Denmark. *Int. J. Cancer*, **119**, 183–191.
- Kanehisa,M. and Goto,S. (2000) KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.*, **28**, 27–30.
- Kanehisa,M. et al. (2006) From genomics to chemical genomics: new developments in KEGG. *Nucleic Acids Res.*, **34**, D354–D357.
- Kanehisa,M. et al. (2008) KEGG for linking genomes to life and the environment. *Nucleic Acids Res.*, **36**, D480–D484.
- Kim,J.W. and Dang,C.V. (2006) Cancer's molecular sweet tooth and the Warburg effect. *Cancer Res.*, **66**, 8927–8930.
- Li,M. et al. (2008) Symbiotic gut microbes modulate human metabolic phenotypes. *Proc. Natl Acad. Sci. USA*, **105**, 2117–2122.
- Matys,V. et al. (2003) TRANSFAC: transcriptional regulation, from patterns to profiles. *Nucleic Acids Res.*, **31**, 374–378.
- Matys,V. et al. (2006) TRANSFAC and its module TRANSCOMP: transcriptional gene regulation in eukaryotes. *Nucleic Acids Res.*, **34**, D108–D110.
- Moreno-Sanchez,R. et al. (2007) Energy metabolism in tumor cells. *FEBS J.*, **274**, 1393–1418.
- Nakamura,R. et al. (2005) Gene expression profiling of dexamethasone-treated RBL-2H3 cells: induction of anti-inflammatory molecules. *Immunol. Lett.*, **98**, 272–279.
- Nam,H. et al. (2009) Computational identification of altered metabolism using gene expression and metabolic pathways. *Biotechnol. Bioeng.*, **103**, 835–843.
- Phillips,M. et al. (2003) Volatile markers of breast cancer in the breath. *Breast J.*, **9**, 184–191.
- Phillips,M. et al. (2006) Prediction of breast cancer using volatile biomarkers in the breath. *Breast Cancer Res. Treat.*, **99**, 19–21.
- Poulter,J.M. et al. (1985) Tryptophan metabolism in patients with breast cancer. *Acta Vitaminol. Enzymol.*, **7**, 93–97.
- Sigoillot,F.D. et al. (2004) Breakdown of the regulatory control of pyrimidine biosynthesis in human breast cancer cells. *Internat. J. Cancer*, **109**, 491–498.
- Sreekumar,A. et al. (2009) Metabolomic profiles delineate potential role for sarcosine in prostate cancer progression. *Nature*, **457**, 910–914.

- Stein, S.E. (1999) An integrated method for spectrum extraction and compound identification from gas chromatography/mass spectrometry data. *J. Am. Soc. Mass Spectrom.*, **10**, 770–781.
- Subramanian, A. *et al.* (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA*, **102**, 15545–15550.
- Wishart, D.S. *et al.* (2007) HMDB: the human metabolome database. *Nucleic Acids Res.*, **35**, D521–D526.
- Yeh, C.S. *et al.* (2006) Fatty acid metabolism pathway play an important role in carcinogenesis of human colorectal cancers by Microarray-Bioinformatics analysis. *Cancer Lett.*, **233**, 297–308.