Discovery of Urinary Biomarkers in Patients with Breast Cancer Based on Metabolomics

Jeongae Lee[1,†], Han Min Woo[2,†], Gu Kong[3], Seok Jin Nam[4], and Bong Chul Chung[1,*]

[1]*Molecular Recognition Research Center,Korea Institute of Science and Technology, Hawolgok-dong 39-1, Seongbuk-gu, Seoul 136-791, Korea*

[2]*Clean Energy Research Center,Korea Institute of Science and Technology, Hawolgok-dong 39-1, Seongbuk-gu, Seoul 136-791, Korea*

[3]*Department of Pathology, School of Medicine, Hanyang University, Seoul 133-792, Korea*

[4]*Division of Breast-Endocrine Surgery, Department of Surgery, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul 135-710, Korea*

Abstract: A metabolomics study was conducted to identify urinary biomarkers for breast cancer, using gas chromatography-mass spectrometry (GC-MS) and liquid chromatography-mass spectrometry (LC-MS), analyzed by principal components analysis (PCA) as well as a partial least squares-discriminant analysis (PLS-DA) for a metabolic pattern analysis. To find potential biomarkers, urine samples were collected from before- and after-mastectomy of breast cancer patients and healthy controls. Androgens, corticoids, estrogens, nucleosides, and polyols were quantitatively measured and urinary metabolic profiles were constructed through PCA and PLS-DA. The possible biomarkers were discriminated from quantified targeted metabolites with a metabolic pattern analysis and subsequent screening. We identified two biomarkers for breast cancer in urine, β-cortol and 5-methyl-2-deoxycytidine, which were categorized at significant levels in a student t-test (p-value < 0.05). The concentrations of these metabolites in breast cancer patients significantly increased relative to those of controls and patients after mastectomy. Biomarkers identified in this study were highly related to metabolites causing oxidative DNA damage in the endogenous metabolism. These biomarkers are not only useful for diagnostics and patient stratification but can be mapped on a biochemical chart to identify the corresponding enzyme for target identification via metabolomics.

Key words: Metabolomics, Breast Cancer, Biomarker, PLS-DA, Urine

Introduction

Breast cancer is a malignant tumor that starts from cells of the breast and is a common cancer in females. The incidence of breast cancer in Korea has increased three-fold since 1983. It is the second leading cancer causing mortal death of Korean women after stomach cancer.[1] Although the causes of breast cancer are still unknown, there are many risk factors associated with breast cancer. An inherited gene abnormality is associated with a relatively higher risk for developing the disease.[2] Prolonged, uninterrupted exposure to estrogen can increase breast cancer risk and both normal and abnormal breast cell growth are stimulated by the presence of estrogen.[3] Additionally, smoking, diet, and stress have been associated with the elevation of breast cancer risk.[4] Fortunately, if the disease is discovered early, patients have more treatment options and a better chance for long-term recovery. Many researchers have accordingly tried to find early diagnostic biomarkers using various omics tools.

Biomarker discovery is a non-invasive method to find relevant markers rapidly without detailed insight into mechanisms of a given disease. By screening many possible biomolecules simultaneously, techniques based on genomics and proteomics have been used to pursue high-throughput screening. However, genomics and proteomics assignments are not specific enough to describe biochemical functions and underlying biological roles. A comprehensive analysis in which all metabolites of a biological system are identified and quantified is needed.

Such an approach is called metabolomics.[5] Metabolomics is the characterization of the function of genes and cellular responses in cells, organs, and biological fluids by profiling changes in metabolites generated under specific conditions. The goals of this field are to find the alteration of metabolites in disease, develop a new drugs, and prevent disease and provide early diagnosis.[6] Metabolic profiling, one of the tools of metabolomics, offers a means to detect and characterize a large number of substances simultaneously, find new and unexpected substances, determine the changes in the ratio of different components, and give precise information, for instance, on steroid hormones or organic acids, compared to a single test or a group test.

In this study, we used urine samples from clinically diagnosed breast cancer patients and healthy volunteers for comparison and identified novel urinary biomarkers for breast cancer. Metabolic profiling of endogenous steroids (androgens, estrogens and corticoids), nucleosides, and polyols that elevate breast cancer risk was performed by gas chromatography-/liquid chromatography-mass spectrometry, and quantified data were subsequently analyzed by a multivariate data analysis. Although most metabolomics studies employed mice or cell lines, there have been few reports involving human urine, especially using mass spectrometry. The intensity of any metabolite can directly be compared to the intensity of the same metabolite in another sample. The difference may be characteristic of a specific disease. Mass spectrometry coupled with a chromatographic technique has been utilized to quantify physiological compounds of metabolic pathways. Data collected from patients with breast cancer and controls were process and then analyzed by a partial least squares discriminant analysis (PLS-DA) to discover urinary biomarkers.

Experimental

Patients and sample collection

Patients with a diagnosis of invasive breast cancer (n = 6, age = 43.83 ± 7.28) and subjects with no evidence of invasive breast cancer as normal controls (n = 11, age = 49.75 ± 6.85) visited Samsung Hospital and College of Medicine, Hanyang University, Seoul, Korea. The study protocol was approved by the ethical committee of Hanyang University and Samsung Medical Center, Seoul, Korea, and informed consent was obtained from each subject. All cases and controls in this study underwent the same diagnostic procedures, i.e., breast physical examination, mammography, and ultrasonography according to the American Joint Committee on Cancer staging system.[7] Breast cancer patients were predominantly stage IIa. Most urine samples for patients were collected within a week before mastectomy and within a month after mastectomy. Urine samples from patients and controls were stored at −20°C in a freezer for further analysis. The urinary creatinine value was measured by the Jaffè method.

GC-MS and LC-MS analysis

Endogenous steroids (12 androgens, 12 corticoids, and 13 estrogens) in urine samples of breast cancer patients and controls were analyzed by Hewlett-Packard GC-MS systems consisting of a model 5890A GC, 5970B mass-selective detector, and a HP 59970C MS Chemstation.[8,9] Fourteen of nucleosides in urine samples were analyzed by a column-switching HPLC system using a Shiseido Nanospace SI-2 HPLC system model 3001 (Shiseido Co., Tokyo, Japan) equipped with an auto injector (200 mL loop; model 3023) and a dual, six-way switching value unit (model 3012).[10] Additionally, urine samples for the quantification of polyols were analyzed by a simultaneous measurement method using GC-MS.[11] Names and ID numbers of targeted metabolites are listed in Table 1. Standard mixtures of each targeted metabolite were first analyzed to obtain a linear calibration curve to quantify targeted metabolites. The linearity of each calibration curve hit the high regression coefficients.

Sample preparation

Endogenous steroids in urine samples were extracted with an organic solvent after enzymatic hydrolysis, and polyols were diluted with distilled water. The extracted residues were derivatized by a previously described method.[8,9,11] The nucleosides in the urine sample were centrifuged, filtered, and directly analyzed by column-switching HPLC-positive ionization tandem mass spectrometry as described previously.[10]

Statistical analysis

For data acquisition, G170AA Chemstation version A.03.00 and Xcalibur 2.0 SR2 were used for analysis of GC-MS and LC-MS, respectively. To identify targeted metabolites, several chromatograms were evaluated using selective mass ion based on the retention time and m/z (ratio of mass to charge). Using script macro programming performed on Chemstation software, the integrated area of a selected ion mass that is able to quantify the targeted metabolite was calculated. Each calculated area was divided by the area of an internal standard in a single run to compensate analytical errors in the same batch. The relative area of the targeted metabolites was used to calculate their quantitative concentrations using the linear calibration curve. The micro molar concentration (μmol/L) of the targeted metabolites was the calibrated by the gram of urinary creatinine concentration. When data were missing for detection, it was considered as a blank for further treatment of missing value. All data manipulation was performed by Matlab version 7.0.4 R14 (The Mathworks Inc.) and an Excel spread sheet (Microsoft Inc. WA). Each metabolite was statistically evaluated by a student t-test ($p < 0.05$).

Data acquired from a study of metabolomics showed a high dimension of measured parameter and a low dimension of designed experiments, which characterize

Table 1. Targeted Metabolites Analyzed by GC-MS and LC-MS

Name of metabolites (ID number)	Selected mass ions (m/z)	RRT* (min)	Name of metabolites (ID number)	Selected mass ions (m/z)	RRT (min)
Androgens			2-Hyroxyestradiol-3-methylether (63)	446, 315, 416	1.20
Androsterone (1)	272[†], 347, 362	0.392	2-Hydroxyestrone (61)	430, 415, 345	1.22
Etiocholanolone (2)	272, 347, 362	0.407	16α-Hyroxyestrone (64)	286, 430, 415	1.22
Dehydroepiandrosterone (3)	304, 360, 129	0.441	2-Methoxyestradiol (65)	446, 315, 416	1.23
4-Androstene-3,17-dione (6)	358, 286, 244	0.445	16-Ketoestradiol (66)	430, 285, 245	1.30
11-Ketoandrosterone (7)	376, 286, 271	0.449	17-Epiestriol (67)	504, 311, 345	1.34
11-Ketoetiocholanolone (8)	271, 376, 286	0.454	Estriol (68)	504, 311, 345	1.38
Testosterone (10)	432, 360, 417	0.507	16-Epiestriol (69)	504, 311, 345	1.44
5-Androstene-3α,17β-diol (9)	239, 434, 344	0.508	**Nucleosides[‡]**		
Dihydrotestosterone (4)	434, 405, 143	0.513	Pseudouridine (25)	245	1.74
11β-Hydroxyandrosterone (11)	199, 522, 450	0.544	Cytidine (26)	244	2.38
16α-Hydroxydehydroepiandrosterone (5)	214, 343, 304	0.590	Uridine (27)	245	3.53
11β-Hydorxyetiocholanolone (12)	156, 522, 360	0.561	1-Methyl-adenosine (28)	282	4.11
Corticoids			5-Methyl-cytidine (29)	258	4.70
Tetrahydro-11-deoxycorticosterone (15)	375, 463, 257	0.733	5-Methyl-2'deoxycytidine (30)	259	5.5.
Tetrahydro-11-deoxycortisol (14)	435, 345, 255	0.735	5-Hydroxymethyl-2-deoxyuridine (36)	242	5.56
Tetrahydro-11-dehydrocorticosterone (13)	348, 492, 402	0.786	Guanosine (31)	242	5.70
Tetrahydrocortisone (16)	449, 359, 550	0.795	3-Methyluridine (32)	259	5.76
Tetrahydrocortisol (17)	523, 343, 253	0.808	Adenosine (33)	268	5.82
5α-Tetrahydrocortisol (18)	253, 523, 343	0.813	2-Deoxyguanosine (37)	268	5.86
α-Cortolone (19)	449, 359, 369	0.816	8-Hydroxy-2-deoxyguanosine (38)	284	5.91
Tetrahydrocorticosterone (20)	283, 463, 255	0.837	N,N-dimethylguanosine (34)	312	6.11
β-Cortol (22)	253, 445, 343	0.840	5-Deoxyadenosine (35)	252	6.24
α-Cortol (23)	253, 523, 343	0.865	**Polyols**		
5α-Tetrahydrocorticosterone (24)	230, 638, 147	0.949	Adonitol (49)	303	0.72
β-Cortolone (21)	449, 359, 269	1.839	Arabitol (50)	303	0.73
Estrogens			Xylitol (51)	303	0.79
Estrone (57)	342, 257, 218	0.93	Glucose (52)	169	1.00
17β-Estradiol (58)	416, 285, 232	1.00	Mannitol (53)	171	1.07
2-Meoxyestrone (59)	372, 342, 357	1.15	Dulcitol (54)	375	1.10
4-Methoxyestradiol (60)	446, 315, 325	1.17	Sorbitol (55)	375	1.11
2-Methoxyestradiol-3-methylether (62)	388, 257, 269	1.18	Myo-Inositol (56)	375	1.12

*RRT stands for relative retention time, which was adjusted by corresponding internal standards.
[†]Quantitative ions are underlined.
[‡]For analysis of nucleoside the collision-induced dissociation of $[M+H]^+$ was characterized in positive mode using LC-MS.

multivariate data in a single batch experiment. This is not well described to find the statistical results by a univariate data analysis such as the student t-test or ANOVA due to the lack of consideration of cross-data implementation. Therefore, for analysis of a metabolomic study, multivariate data analyses such as the principle component analysis (PCA) as an unsupervised method and partial least square–discriminant analysis (PLS-DA) as a supervised method are more adaptable to classified disease groups.[12,13] PLS-DA maximizes the covariance between the predicting data sets,

the concentration of targeted metabolites and the class assignment, respectively. In this study, we used SIMCA software (Umetrics Inc. Sweden) to analyze metabolite profiles for breast cancer patients and normal controls. Urinary concentrations normalized by gram of creatinine concentration were scaled and centered prior to PCA and PLS-DA. For the analysis with the supervised method, PLS-DA, the statuses of breast cancer patients and normal controls were set to individual classes for the prediction of modeling and visualization of results.

Table 2. Quantification of statistically significant targeted metabolites

Target metabolites ID	Average concentration (Mean: μmol/g of creatinine ± SD)			p-value(<0.05) Control vs. Before / Before vs. After
	controls	before-mastectomy	after-mastectomy	
Androgens				
Androsterone (1)	3.59 ± 2.29	1.98 ± 0.68	2.69 ± 0.86	0.050 / 0.154
Etiocholanolone (2)	0.59 ± 0.42	0.47 ± 0.18	0.57 ± 0.22	0.424 / 0.602
Dehydroepiandrosterone (3)	2.48 ± 2.02	0.96 ± 0.63	1.60 ± 0.55	0.039 / 0.173
Dihydrotestosterone (4)	4.93 ± 4.33	0.95 ± 0.53	1.98 ± 0.68	0.012 / 0.042
16α-Hydroxydehydroepiandrosterone (5)	3.10 ± 2.11	1.51 ± 0.52	2.56 ± 0.82	0.035 / 0.040
4-Androstene-3,17-dione (6)	3.06 ± 1.53	2.48 ± 0.80	2.62 ± 1.32	0.322 / 0.718
11-Ketoandrosterone (7)	0.64 ± 0.28	0.36 ± 0.16	0.54 ± 0.33	0.019 / 0.314
11-Ketoetiocholanolone (8)	0.44 ± 0.30	0.41 ± 0.16	0.57 ± 0.29	0.746 / 0.419
5-Androstene-3α,17β-diol (9)	1.15 ± 0.73	0.87 ± 0.57	0.97 ± 0.77	0.398 / 0.963
Testosterone (10)	4.84 ± 4.01	8.79 ± 5.60	4.85 ± 3.81	0.166 / 0.068
11β-Hydroxyandrosterone (11)	2.46 ± 1.40	1.73 ± 0.59	2.10 ± 1.03	0.156 / 0.203
11β-Hydorxyetiocholanolone (12)	1.17 ± 0.58	1.38 ± 0.47	1.47 ± 0.74	0.434 / 0.782
Corticoids				
Tetrahydro-11-dehydrocorticosterone (13)	9.02 ± 3.70	5.19 ± 2.24	4.85 ± 2.78	0.018 / 0.759
Tetrahydro-11-deoxycortisol (14)	10.34 ± 6.82	3.91 ± 1.92	8.44 ± 3.56	0.012 / 0.053
Tetrahydro-11-deoxycorticosterone (15)	6.95 ± 7.25	7.96 ± 3.62	10.22 ± 8.05	0.707 / 0.634
Tetrahydrocortisone (16)	34.57 ± 24.06	35.14 ± 27.85	25.79 ± 16.11	0.967 / 0.385
Tetrahydrocortisol (17)	14.95 ± 10.40	15.20 ± 12.04	11.15 ± 6.96	0.967 / 0.385
5α-Tetrahydrocortisol (18)	18.09 ± 12.59	18.39 ± 14.57	13.49 ± 8.43	0.967 / 0.385
α-Cortolone (19)	15.08 ± 11.46	7.09 ± 4.06	18.63 ± 9.42	0.056 / 0.032
Tetrahydrocorticosterone (20)	7.63 ± 4.28	5.55 ± 2.39	5.18 ± 2.97	0.219 / 0.759
β-Cortolone (21)	17.42 ± 8.92	15.28 ± 7.89	22.17 ± 8.59	0.620 / 0.256
β-Cortol (22)	2.97 ± 1.11	6.17 ± 2.81	3.09 ± 1.03	0.037 / 0.013
α-Cortol (23)	5.64 ± 2.77	4.93 ± 2.60	5.17 ± 2.49	0.608 / 0.930
5α-Tetrahydrocorticosterone (24)	5.07 ± 2.52	8.92 ± 6.48	5.81 ± 3.06	0.213 / 0.297
Nucleosides				
Pseudouridine (25)	130.34 ± 77.81	84.57 ± 52.53	77.24 ± 56.17	0.172 / 0.570
Cytidine (26)	0.91 ± 0.79	1.39 ± 1.00	0.91 ± 1.21	0.341 / 0.330
Uridine (27)	1.36 ± 1.16	2.18 ± 1.11	1.40 ± 0.91	0.181 / 0.263
1-Methyl-adenosine (28)	14.82 ± 10.30	27.26 ± 13.09	13.62 ± 18.55	0.077 / 0.215
5-Methyl-cytidine (29)	71.59 ± 67.57	97.19 ± 20.31	55.16 ± 49.03	0.266 / 0.122
5-Methyl-2'deoxycytidine (30)	1.51 ± 1.17	1.48 ± 0.69	0.97 ± 0.60	0.950 / 0.249
Guanosine (31)	0.52 ± 0.51	1.68 ± 2.26	0.61 ± 0.50	0.266 / 0.277
3-Methyluridine (32)	2.20 ± 1.46	3.48 ± 2.41	2.01 ± 1.53	0.272 / 0.246
Adenosine (33)	2.81 ± 3.55	4.17 ± 6.65	2.16 ± 2.01	0.656 / 0.488
N,N-dimethylguanosine (34)	1.33 ± 1.92	4.07 ± 2.41	2.94 ± 3.13	0.042 / 0.498
5-Deoxyadenosine (35)	0.74 ± 0.65	0.45 ± 0.54	0.64 ± 0.60	0.354 / 0.690
5-Hydroxymethyl-2-deoxyuridine (36)	0.09 ± 0.06	0.29 ± 0.23	0.07 ± 0.06	0.095 / 0.013

Table 2. Continued

Target metabolites ID	Average concentration (Mean: μmol/g of creatinine ± SD)[§]			p-value(<0.05) Control vs. Before / Before vs. After
	controls	before-mastectomy	after-mastectomy	
2-Deoxyguanosine (37)	0.05 ± 0.07	0.00 ± 0.00	0.33 ± 0.38	0.042 / 0.089
8-Hydroxy-2-deoxyguanosine (38)	0.07 ± 0.06	0.27 ± 0.19	0.11 ± 0.07	0.048 / 0.146
Polyols				
Adonitol (49)	101.75 ± 52.45	53.70 ± 14.85	42.50 ± 10.32	0.014 / 0.181
Arabitol (50)	688.55 ± 403.60	258.72 ± 116.96	193.13 ± 96.55	0.006 / 0.168
Xylitol (51)	99.27 ± 62.72	42.71 ± 13.12	33.11 ± 7.83	0.014 / 0.287
Glucose (52)	662.52 ± 324.29	258.87 ± 76.09	172.13 ± 37.36	0.002 / 0.097
Mannitol (53)	3129.42 ±	429.12 ± 510.69	610.11 ± 644.44	0.242 / 0.767
Dulcitol (54)	43.00 ± 21.74	41.72 ± 27.00	35.90 ± 21.15	0.923 / 0.413
Sorbitol (55)	99.34 ± 65.76	50.53 ± 23.46	40.76 ± 20.49	0.044 / 0.301
Myo-Inositol (56)	199.29 ± 135.64	95.44 ± 74.08	154.62 ± 210.52	0.059 / 0.612
Estrogens				
Estrone (57)	0.03 ± 0.04	0.02 ± 0.02	0.03 ± 0.02	0.576 / 0.732
17β-Estradiol (58)	0.01 ± 0.01	0.00 ± 0.00	0.00 ± 0.00	0.147 / 0.796
2-Meoxyestrone (59)	0.01 ± 0.01	0.00 ± 0.00	0.01 ± 0.01	0.048 / 0.427
4-Methoxyestradiol (60)	0.01 ± 0.01	0.00 ± 0.00	0.00 ± 0.00	0.472 / 0.434
2-Hydroxyestrone (61)	0.01 ± 0.02	0.01 ± 0.01	0.01 ± 0.01	0.881 / 0.901
2-Methoxyestradiol-3-methylether (62)	0.03 ± 0.02	0.01 ± 0.01	0.01 ± 0.01	0.049 / 0.705
2-Hyroxyestradiol-3-methylether (63)	0.00 ± 0.00	0.00 ± 0.00	0.01 ± 0.00	0.474 / 0.057
16α-Hyroxyestrone (64)	0.01 ± 0.01	0.01 ± 0.01	0.01 ± 0.01	0.353 / 0.726
2-Methoxyestradiol (65)	0.01 ± 0.02	0.00 ± 0.00	0.00 ± 0.00	0.059 / 0.003
16-Ketoestradiol (66)	0.01 ± 0.01	0.10 ± 0.10	0.01 ± 0.01	0.109 / 0.076
17-Epiestriol (67)	0.05 ± 0.10	0.00 ± 0.01	0.01 ± 0.01	0.135 / 0.722
Estriol (68)	0.02 ± 0.03	0.04 ± 0.04	0.03 ± 0.02	0.407 / 0.918
16-Epiestriol (69)	0.01 ± 0.03	0.03 ± 0.05	0.00 ± 0.01	0.561 / 0.316

[§]SD stands for standard deviation. Underline (p-value < 0.05)

Results

Quantification of targeted metabolites using GC-MS and LC-MS

For the metabolic profiling of endogenous steroids, which are risk factors for breast cancer, 11 androgens, 11 corticoids, and 13 estrogens in urinary samples were quantified in our previous works.[8,9] In addition, 14 nucleosides and 8 polyols were also analyzed by previously described the methods.[10,11] The concentrations of metabolites were quantified by their linear regression curve, resulting in good linearity with a regression coefficient of over 0.99. Data from sets of control vs. before-mastectomy and before- vs. after-mastectomy are shown in Table 2. The concentrations of 15 targeted metabolites among these targeted metabolites for the control group and patients changed statistically significant by a student t-test ($p < 0.05$). In particular, the concentrations of nucleosides in patients before-mastectomy increased significantly relative to those of the control group and after-mastectomy, whereas the concentrations of estrogens, corticoids, and androgens decreased. The metabolic profiles of polyols decreased significantly in the after-mastectomy group compared to those in the control and before-mastectomy groups. All quantified metabolites were conducted on the Excel® molar spreadsheet to normalize concentrations to the gram of creatinine concentration.

Multivariate data analysis

Subjected to an inter-group analysis, normalized data of controls and patients (before- and after-mastectomy) were combined into a single datasheet for PCA and PLS-DA. Using SIMCA software, PCA as an unsupervised method and PLS-DA as a supervised method were conducted to discriminate metabolic profiles from urinary samples (Figure 1). The metabolic pattern obtained using PLS-DA provided clear classification between the control group and the patients group than that obtained with PCA (figure not shown). Among the results of a multivariate analysis using each of

Figure 1. (A) PLS-DS score plot of targeted metabolites in breast cancer patient's urine (blue circle symbol represents control group; black triangle before-mastectomy group; red square after-mastectomy group). The number in the after-mastectomy group denotes the collection period (day) after mastectomy, (B) PLS-DA loading plot of targeted metabolites in breast cancer patient's urine (each black triangle represents targeted metabolite with its ID number on the top of the symbol, and (C) Variable Importance (VIP value) in the projection to reflect the importance of selected metabolites in the model. VIP values are related to the correlation of the metabolites.

these methods, PLS-DA showed a better result with checking by the fraction of the variation of the variable $Q^2VY = 0.32$ as estimated by cross validation than the results of PCA (figure not shown). The PLS-DA score plot for the control groups and breast cancer patients is illustrated in Figure 1A. The first component t[1] (the fraction of the variation of the response $R^2VY = 0.44$) and the second component t[2] ($R^2VY = 0.36$) of the score plot were fully able to discriminate before-mastectomy patients from controls. Targeted metabolic profiles in the urine of patient's before-mastectomy were displayed in the third quarter while those of patient's after-mastectomy were shown in the second quarter. Metabolic profiles for the control group were distributed around the positive t[1] axis. Also, we found that the metabolic patterns from the state before-mastectomy were restored after-mastectomy relative to those of the controls. As shown in Figure 1A, the later urine for the after-mastectomy group was collected, the greater the similarity of the metabolic pattern to that of the control.

Identification and validation of biomarkers

We defined the pool of targeted metabolites in endocrinal metabolism and investigated the correlation between the metabolic pattern and the targeted metabolites to discover possible biomarkers for breast cancer in urine. Using targeted metabolomics, we selected several possible urinary biomarkers for breast cancer by interpreting the geometrical relation and statistical analysis of both the PLS-DA score plot and loading plots (Figures 1A and B). We were interested in particular metabolites in the region of the before-mastectomy patient group (3rd quarter) that can be regarded as biomarkers for breast cancer in urine. Sixteen possible metabolites were selected from the first geometric link (Table 3). We evaluated their possibilities as potential biomarkers on the basis of statistical significance. First, we listed possible metabolites by the second criterion, Variable Importance in the projection values (VIP), which reflect the importance of variables both with respect to samples (Y) and metabolites (X). Among possible biomarkers, targeted metabolites were screened by one term to other with larger VIP, larger than 1 due to the average VIP equal to 1 (Figure 1C). After finding most relevant metabolites by VIP, we compared possible biomarkers with statistical results of quantification of targeted metabolites (Table 2). As a result of the selection process, two metabolites, β-cortol and 5-hydroxymethyl-2-deoxyuridine, were chosen as potential urinary biomarkers for breast cancer (Figure 2). The metabolic profiles of these potential biomarkers for breast cancer selected by using metabolomics were distinct from those of the before-mastectomy group, showing a decrease after the mastectomy, and those of the after-mastectomy group were restored to the level of the control group, which can be considered as the normal condition. Those potential metabolites were qualified by student t-test ($p < 0.05$), compared to the control group.
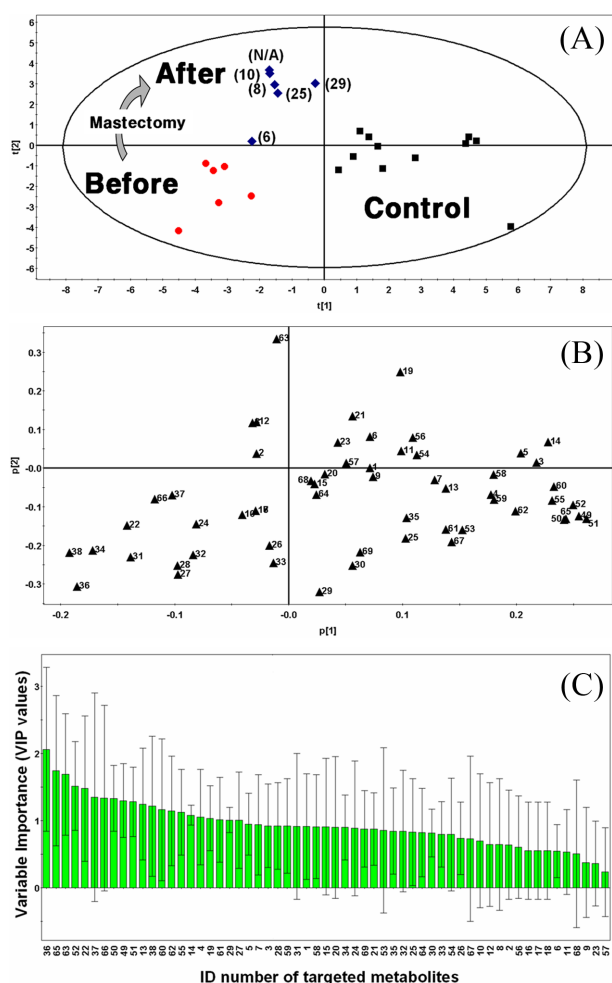
Table 3. Variable Importance in the Projection for Targeted Metabolites Selected in the 3rd Quarter of PLS-DA Loading Plot

| Selected Target Metabolite (Metabolite ID)	Variable Importance	Selected Target Metabolite (Metabolite ID)	Variable Importance
5-Hydroxymethyl-2-deoxyuridine (36)[††]	2.06	Guanosine (31)	0.92
β-Cortol (22)[††]	1.48	5α-Tetrahydrocorticosterone (24)	0.89
2-deoxy-guanosine (37)	1.35	3-Methyluridine (32)	0.84
16-Ketoestradiol (66)	1.34	Adenosie (33)	0.80
8-Hydroxy-2-deoxyguanosine(38)	1.21	Cytidine (26)	0.74
Uridine (27)	1.01	Testosterone (10)	0.70
1-Methyl-adenosine (28)	0.92	5α-Tetrahydrocortisol (18)	0.55
N,N-dimethylguanosine (34)	0.90	Tetrahydrocortisol (17)	0.55

[††]Marked metabolites (p-value < 0.05)
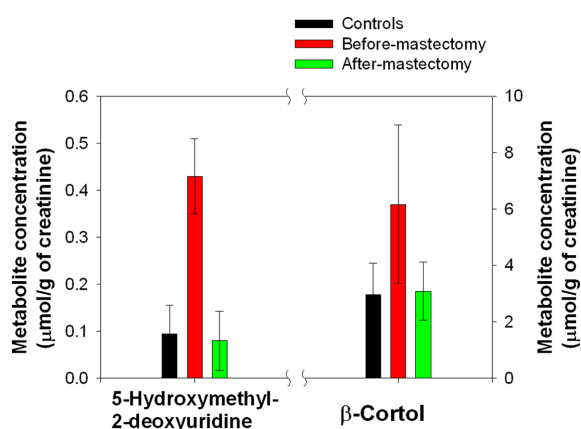
Figure 2. Metabolite profiles of potential biomarkers for breast cancer. Black bar with error bar (s.d.) stands for control group, red bar before-mastectomy, and green bar after-mastectomy. The proposed potential biomarkers for breast cancer selected from a metabolomic approach showed significant changes of the concentration (student t-test; P < 0.05), increasing in breast cancer patients and then decreasing after mastectomy.

For validation of the PLS-DA model, the fraction of the variable predicted by a component was calculated by estimation of cross-validation (Q2 = 0.32). No components of PLS-DA overrode the Hotelling T2 range by 17.04 T2Crit (99%). Moreover, the order of samples was randomly permuted 20 times and all the permuted samples fit the original matrix samples. There was little discrepancy between the observed and predicted samples due to human variation.

Discussion

In this paper we suggested two potential biomarkers for breast cancer, 5-hydroxymethyl-2-deoxyuridine and β-cortol. 5-Hydroxymethyl-2-deoxyuridine has been reported elsewhere to be an oxidative DNA damage marker. The formation of hydroxylated DNA bases is an important form of oxidative DNA damage. Oxidative DNA damage can result from numerous endogenous metabolic processes as well as from exposure to environmental and dietary oxidants. In recent

Table 4. Average Concentrations of 5-hydroxymethyl-2-deoxyuridine in Urine Categorized by Cancer Stages

Breast Cancer Stage	≤ I	I	IIa	IIb	IIIa	IIIc
Before-mastectomy	0.64 ± 0.12	0.44 ± 0.21	0.43 ± 0.20	0.37 ± 0.20	0.41 ± 0.27	0.32 ± 0.10
After-mastectomy	0.13 ± 0.12	0.10 ± 0.09	0.11 ± 0.07	0.08 ± 0.04	0.11 ± 0.06	0.09 ± 0.14
Control	0.07 ± 0.05					

Unit: μmol/g of Creatinine; mean ± Standard Deviation
The urine samples of patients with breast cancer ($n = 140$; 46.6 ± 7.8 years) and normal female subjects ($n = 120$; 47.2 ± 7.8 years) were analyzed directly by column-switching LC using a pre-column for clean-up and an analytical column for the complete separation and tandem mass spectrometry with positive electrospray ionization as described previously [10].

years, it has been reported that oxidative DNA damage plays an important role in numerous disease processes implicated in carcinogenesis and neurodegenerative diseases.[14,15] Djuric et al. indicated that 5-hydroxymethyl-2-deoxyuridine was significantly increased as a marker in DNA from the blood of breast cancer patients.[16] We also reported that there were increased levels of urinary nucleosides in breast cancer patients,[10] and we analyzed the level of 5-hydroxymethyl-2-deoxyuridine according to breast cancer staging. The results showed the highest level below stage I compared to the other stages (Table 4), which was consistent with levels of 5-hydroxymethyl-2-deoxyuridine when oxidative DNA damage is elevated in human sera, implicating early signal of cancer risk in benign breast cancer and healthy women who have a family history of breast cancer.[17] This means that it indirectly supports the predictive validity of our research. Therefore, if an organism (or cell-line) is subjected to either biological or chemical stress, the specific biochemical abnormalities caused from the direct result of that stressor could be apparent. From the trace of metabolic patterns, the demonstration of these changes provides a unique way to understand pathological statesbecause of the ability to directly relate the data in the context of normal biochemistry. Moreover, the level of 5-hydroxymethyl-

Jeongae Lee, Han Min Woo, Gu Kong, Seok Jin Nam, and Bong Chul Chung

2-deoxyuridine for the after-mastectomy group decreased significantly ($p < 0.001$) according to the day of urine collection, from day 6 to day 29, and become closer to that of the control group (Figure 1A). This may indicate recovery from DNA damage.

Another suggested potential biomarker is the corticosteroid β-cortol. Most corticosteroids decreased in patients with breast cancer, whereas β-cortol significantly increased ($p<0.037$), as shown in Table 2. Seraj et al. reported that corticosteroids with a carbonyl group at carbon seventeen (C17) of the cholesterol skeleton formed in the formation of a covalent bond with DNA adducts in vitro.[18] β-Cortol, therefore, probably fails to form a DNA adduct since it does not have the carbonyl group at C17 of the cholesterol skeleton, and hence it could be responsible for increased urinary excretion of β-cortol. In the case of other women's diseases such as polycystic ovary syndrome and uterine leiomyomas, β-cortol was also raised in the disease condition.[19,20] These results agree with our findings and it could be speculated that, as a biomarker candidate, the urinary level of β-cortol may be related to hormone dependence in women's diseases. The validity of β-cortol as a new biomarker for breast cancer requires further investigation on the molecular level.

Conclusion

We presented the first report of urinary biomarkers in breast cancer patients based on metabolomics. It is suggested that the urinary metabolic profiles from mass spectrometry combined with PLS-DA can be used not only to differentiate breast cancer patients from controls, but also to discover and identify potential biomarkers. Metabolomics provides a more comprehensive look at the complex changes under way in hundreds of molecules as a disease begins to develop. Unlike proteomics or genomics, where the molecules are in one class, either proteins or nucleic acids, the metabolome comprises almost all classes of compounds that are found in tissues. Thus metabolomic approaches can be used to discover and further develop metabolic biomarkers that are distinct in tumor cell environment or in pre-cancerous tissues and may be applicable to personalized care and medicine.

Acknowledgements

Author's contributions

JL and HMW contributed equally to this work, performed the statistical analysis with interpretation, and drafted the manuscript. JL carried out GC-MS and LC-MS analyses. HMW carried out the collection of the data and the statistical analysis. GK and SJN performed pathologic examination and the collection of urine samples with patients and controls. BCC and GK conceived the study and participated in its design and coordination. BCC approved the final manuscript. All authors read and approved the final version of the manuscript.

References

1. The Korean Ministry of Health and Welfare: Cancer Registry Reports. *Seoul: The Korean Ministry of Health and Welfare* **2002**.
2. Chaudru, V.; Chompret, A.; Bressac-de Paillerets, B.; Spatz, A; Avril, M. F.; Demenais, F. *J. Natl. Cancer Inst.* **2004**, 96, 785.
3. Eiassen, A. H.; Missmer, S. A.; Tworoger, S. S.; Spiegelman, D.; Barbieri, R. L.; Dowsett, M.; Hankinson, S. E. *J. Natl. Cancer Inst.* **2006**, 98, 1406 .
4. Hilakivi-Clarke, L.; Olive, S. E.; Shajahan, A.; Khan, G.; Zhu, Y.; Zwart, A.; Cho, E.; Clarke, R. *J. Nutr.* **2005**, 135, 2946S.
5. Fiehn, O. *Plant Mol. Biol.* **2002**, 48, 155.
6. Brown, M.; Duun, W. B.; Ellis, D. I.; Goodacre, R.; Handl, J.; Knowles, J. D.; O'Hagan, S.; Spasiæ, I.; Kell, D. B. *Metabolomics* **2005**, 1, 39.
7. Singletary, S. E.; Allred, C.; Ashley, P.; Bassett, L. W.; Berry, D.; Bland, K. I.; Borgen, P. I.; Clark, G; Edge, S. B.; Hayes, D. F.; Hughes, L. L.; Hutter, R. V. P.; Morrow, M.; Page, D. L.; Recht, A.; Theriault, R. L.; Thor, A.; Weaver, D. L.; Wieand, H. S.; Greene, F. L. *J. Clin. Oncology* **2002**, 20, 3628 .
8. Lee, S. H.; Kim, S. O.; Kwon, S. W.; Chung, B. C. *Clin. Biochem.* **1999**, 32, 375.
9. Lee, S. H.; Kim, S. O.; Lee, H. D.; Chung, B. C. *Cancer Lett.* **1998**, 133, 47.
10. Cho, S. H.; Jung, B. H.; Lee, S. H.; Lee, W. Y.; Kong, G.; Chung, B. C. *Biomed.Chromatogr.* **2006**, 20, 1229.
11. Lee, J.; Chung, B. C. *J. Chromatogr. B* **2006**, 831, 126.
12. Barker, M.; Rayens, W. *J. Chemometrics* **2003**, 17, 166 .
13. Jia, L.; Wang, C.; Kong, H.; Cai, Z.; Xu, G. *Metabolomics* **2006**, 2, 95.
14. Murata, M.; Kurimoto, S.; Kawanishi, S. *Chem. Res.Toxicol.* **2006**, 19, 1379 .
15. Zana, M.; Szécsényi, A.; Czibula, Á.; Bjelik, A.; Juhasz, A.; Rimanoczy, A.; Szabó, K.; Vetró, Á.; Szûcs, P.; Várkonyi, Á.; Pákáski, M.; Boda, K.; Raskó, I.; Janka, Z.; Kálmán, J. *Biochem. Biophys. Res. Commn.* **2006**, 345, 726.
16. Djuric, Z.; Heilbrun, L. K.; Simon, M. S.; Smith, D.; Luongo, D. A.; LoRusso, P. M.; Martino, S. *Cancer* **1996**, 77, 691.
17. Frenkel, K.; Karkoszka, J.; Glassman, T.; Dubin, N.; Toniolo, P.; Taioli, E.; Mooney, L. A.; Kato, I. *Cancer Epidemiol.Biomarkers Prev.* **1998**, 7, 49.
18. Seraj, M. J.; Umemoto, A.; Tanaka, M.; Kahikawa, A.; Hamada, K.; Monden, Y. *Mutat. Res.***1996**, 370, 49.
19. Tsilchorozidou, T.; Honour, J. W.; Conway, G. S. *J. Clin. Endocrinol. Metab.* **2003**, 88, 5907.
20. Jung, B. H.; Bai, S. W.; Chung, B. C. *Int. J. Gynnecol. Obstst.* **2004**, 84, 55.