Altered levels of nucleoside metabolite profiles in urogenital tract cancer measured by capillary electrophoresis

Ewa Szymańska [a], Michał J. Markuszewski [a,b,*], Marcin Markuszewski [c], Roman Kaliszan [a]

[a] *Department of Biopharmaceutics and Pharmacodynamics, Medical University of Gdańsk, Gen. J. Hallera 107, 80-416 Gdańsk, Poland*
[b] *Department of Toxicology, Ludwik Rydygier Collegium Medicum, Nicolaus Copernicus University in Toruń, M. Skłodowskiej-Curie 9, 85-094 Bydgoszcz, Poland*
[c] *Department of Urology, Medical University of Gdańsk, Kliniczna 1a, 80-402 Gdańsk, Poland*

ABSTRACT

Metabolic profiles of nucleosides studied on the level of urine are closely related to the pathophysiological status of the organism. Posttranscriptional modifications of RNA (mostly tRNA) in cell nucleus are responsible for change of nucleoside levels during malignant disease. In this paper, 256 metabolite profiles from 160 urogenital tract cancer patients and 96 healthy controls, composed of 19 nucleosides were collected and studied with the application of such an approach. This approach comprised of the analysis of urine extracts and the investigation of collected nucleoside and modified nucleoside profiles by advanced statistical data processing tools such as principal component analysis (PCA), hierarchical cluster analysis (HCA), K-Nearest Neighbor method (kNN) and partial least squares-discriminant analysis with probabilistic function (p-PLS-DA). It has been shown that alterations of metabolite profiles in cancer diseases are mainly expressed by the fold change of the urine levels of most nucleosides. In addition, observed metabolite-to-metabolite ratios differ in urogenital cancer patients compared to healthy controls. The obtained relationships between urinary nucleoside profiles and the presence of cancer diseases have been evaluated. Discrimination of the cancer patients and the non-cancer healthy subjects is with 76.5% sensitivity and 80.2% specificity. The presented results prove the usefulness of the metabolomic approach in studying urinary nucleoside profiles with high diagnostic potency in urogenital cancer diseases. Profiles of urinary nucleosides might be employed as a reliable and convenient tool in the diagnostics of urogenital tract cancer diseases.

1. Introduction

Part of the difficulty in understanding cancer diseases arises not only from unknowns regarding what triggers its onset and progression (genome, environmental factors), but also due to uncertainty regarding how best to detect early biological signals that are predictive of subsequent phenotypic changes. The enhanced sensitivity of metabolomics to capture subtle changes in the multiple metabolic paths and cellular levels of metabolites holds promise to identify critical determinants of cancer risk and tumor behavior [1]. The identification of proper biomarkers in biological fluids continues to be a major obstacle for developing effective strategies for cancer diagnosis and therapy.

Monitoring fluctuations in certain metabolite levels is a way to detect early stages in carcinogenesis, predict the aggressiveness of a tumor and/or monitor the response to an intervention [2]. Quantitative analysis of metabolomic techniques, combined with advanced bioinformatic approaches, could enable the development of a profile that characterizes early changes of specific cell or organ functions [3]. Thus, metabolomic approaches are likely to be used to advance research in cancer biology by developing biomarkers which will be useful for distinguishing between precancerous and cancerous states. Understanding the metabolomic changes associated with the shifted balance among growth, stasis, apoptosis and differentiation that characterize tumors remains an important undertaking that will advance early detection. These are for example polyamines involved in cell and tissue proliferation [4,5], nucleosides and modified nucleosides involved in the turnover of ribonucleic acids [6] and pteridines involved in the basic immunological response [7]. The mentioned groups of metabolites have already been investigated in several types of malignant diseases such as leukemia, breast cancer, thyroid cancer, uterine cervical cancer and liver cancer diseases [4,8–11].

* Corresponding author at: Department of Biopharmaceutics and Pharmacodynamics, Medical University of Gdańsk, Gen. J. Hallera 107, 80-416 Gdańsk, Poland. Tel.: +48 58 349 3260; fax: +48 58 349 3262.

E-mail addresses: michal.markuszewski@cm.umk.pl, markusz@amg.gda.pl (M.J. Markuszewski).

A challenge in identifying informative metabolic biomarkers is that there are probably few, if any, metabolites that will be consistently predictive across all cancers. However, it is possible that there are limited numbers of metabolomic processes for any given tumor types; thus, following the metabolomic profiles, rather than individual metabolites, may aid the early diagnosis of disease.

It has already been established that levels of urinary nucleosides might be changed in malignant diseases, such as leukemia or liver cancer [6,12]. However, it is still not known in what extent urinary nucleoside profiles alter in urogenital cancer diseases and what is the character and strength of these alterations.

In 1977 Gehrke et al. [13] published the first article about determination of urinary nucleosides. They successfully determined six nucleosides from healthy and cancer urine samples using an affinity chromatography with a bounded boronate acid. Since then different techniques have been applied for identification and quantification of nucleosides using not only high-performance liquid chromatography (HPLC) [14,15], but also capillary electrophoresis (CE) in conjunction with various types of detectors [4,16–18].

Electromigration techniques are assumed as the most rapidly growing analytical techniques that have had a great impact in biomedical research, clinical and forensic practices in the last decade [19]. The electromigration technique is a serious alternative to chromatographic one and has been used in analytical chemistry since early 1980s of the last century. Compared to other sensitive analytical techniques, such as HPLC, that have been extensively used for analysis of biologically active substances in clinical routines, electromigration technique like CE holds a number of distinct advantages such as a very small sample volume (nL) necessary for a single analysis, rapid analysis, high resolution and relatively low costs. CE play an important role in analytical chemistry as a very useful technique for the analysis of numerous endogenous and exogenous substances presented in biological fluids. There is one main restriction of electromigration techniques with photometric detection that is a higher limits of detection of nucleosides and weaker selectivity in case of complex biological samples. However, this restriction is possible to be eliminated by using different types of detection, e.g. mass spectrometry (MS) detection. The use of CE-MS or CE-MS/MS systems became more and more routine in bioanalysis nowadays.

When CE is coupled with MS, there are two main advantages. The first one is that the molecular mass of analyte can be known and simultaneously the nucleosides can be temporarily identified by comparing the molecular mass with data in the literature and confirmed through other technique such as MS/MS. The second advantage is that the coexisting effluents from CE separation can be separated by MS if they have different molecular mass. Thus, when compared with CE, CE-MS is preferred for developing an extensive metabolic profile.

The goal of our studies was to compare and evaluate differences between urinary nucleoside profiles from patients with diagnosed urogenital tract cancer and healthy controls with no diagnosed cancer disease. To fulfill this goal, a large-scale metabolomic study was performed consisting of five consecutive steps [20]. It included profiles of nucleosides and modified nucleosides taken from the urine samples of 256 healthy and cancer patients. High diversity among the studied population of patients ensured different cancer locations (including bladder, prostate, kidney and testis cancer) and different disease malignancy aimed at finding one set of metabolites valid for all urogenital tract cancer diseases. This could be of utmost assistance in the development of a universal and non-invasive diagnostic method applicable to urogenital tract cancer diseases.

2. Materials and methods

Chemicals and preparation of standard solutions Reference standards for 13 nucleosides, i.e. uridine, pseudouridine, cytidine, 5-methyluridine, inosine, N4-acetylcytidine, guanosine, adenosine, N2,N2-dimethylguanosine, N6-methyladenosine, N1-methyladenosine, xanthosine, and 8-bromoguanosine (internal standard) and creatinine were purchased from Sigma–Aldrich (St. Louis, MO, USA) Phosphoric acid, borax (sodium tetraborate decahydrate), sodium dodecyl sulfate 98.5% and ammonium acetate were from Sigma (Sigma–Aldrich, St. Louis, MO, USA). Formic acid came from Lancaster Synthesis UK (Newgate, Lancashire, UK). Sodium hydroxide, methanol and ammonia were obtained from POCH (Gliwice, Poland). The Affi-gel 601, used as the stationary phase for the extraction of nucleosides from urine,was purchased from Bio-Rad (Hercules, CA, USA). Reversed osmosed deionised water for the preparation of the standard solution, the background electrolyte and other solutions were from MiliQ-Plus system (Millipore, Vienna, Austria).

The 10 mM stock solutions of all standards were prepared in deionised water (except the stock solution of guanosine, which was prepared in 0.1 M NaOH), and kept frozen at −34 °C. The working standard solutions were prepared by dilution of the stock solutions with deionised water to concentrations in the range of 5–5000 μM.

2.1. Study population

Urine samples were collected from 96 healthy controls (with no diagnosed cancer disease) and 160 cancer patients from the Department of Urology at the Medical University of Gdańsk, Gdańsk, Poland. The studies were performed following the patients' informed consents in accordance with the principles embodied in the Declaration of Helsinki and the applied procedure was approved by the Ethical Committee of the Medical University of Gdańsk.

The group of healthy controls (age range 19–86; 60 women and 36 men) consisted of participants without diagnosed urogenital cancer diseases or any other malignant or infectious disease at the time of sample collection. The group of cancer patients (age range 21–92; 40 women and 120 men) included subjects with a known diagnosis of bladder cancer (95 patients, which counts for 59.4% of total cancer cases), kidney cancer (32 patients, 20%), prostate cancer (16 patients, 10%), testis cancer (7 patients, 4.4%) and other malignant diseases of the urogenital tract (10 patients, 6.3%). Cancer patients had various stages of malignant diseases and underwent different therapies (surgery, chemotherapy, radiotherapy). In the case of surgical operations, the urine sample was always collected a day before surgery.

2.2. Urinary nucleoside profiling

Urine samples were prepared by a validated procedure consisting of a selective solid phase extraction (SPE) step followed by the freeze-drying of extract solutions [20]. After sample preparation, nucleoside profiles were assayed by an optimized and validated capillary electrophoretic (CE) method [20]. All urine nucleoside profiles were obtained with a Beckman Coulter P/ACE MDQ system (Beckman Instruments, Fullerton, CA, USA), fitted with a diode array UV-absorbance detector (190–600 nm), a temperature-controlled capillary compartment (liquid-cooled) and a temperature-controlled autosampler (air-cooled). The applied electrophoretic conditions were as follows: 100 mM borate, 72.5 mM phosphate, 160 mM SDS, pH 6.7; 25 kV voltage, 30 °C temperature during analysis; injection 5 s × 0.5 psi; capillary: untreated fused silica 70 cm length to detector, 50 μm I.D.

In the measurement design the total set of samples (256) was divided into two subsets: A (149 samples) and B (107 samples).

Each subset was randomly divided into smaller groups (7–11 samples), which were analyzed in subsequent measurement days. The nucleoside profiles of each sample were measured by CE twice and the mean value of the two measurements was used in further data processing. The performance of the CE system was monitored during measurements by the analysis of nucleoside profiles in quality control samples (QC samples were obtained by the pooling of 30 randomly selected urine extracts).

2.3. Data preparation and pretreatment

The levels of 19 metabolites were determined on the basis of the ratio of the areas of metabolite peak and the internal standard. Calibration curves were obtained separately for samples from subsets A and B. The relative concentrations of unidentified metabolites (seven metabolites included in the profiles) were calculated by use of calibration curves obtained for metabolites with similar migration times. The urinary nucleoside profiles obtained were normalized with the urinary creatinine level measured for each person. The missing values were replaced by the minimal concentration observed.

2.4. Univariate analysis of urinary nucleoside profiles

Statistical calculations were performed with the tools and tests available within STATISTICA 8.0 software (Statsoft Inc., Tulsa, OK, USA). Alterations between the profiles obtained for different groups of patients (cancer patients vs. controls, less than 40-year-old controls vs. more than 40-year-old controls and women controls vs. men controls) were evaluated by the comparison of ranges, medians and mean values of metabolite levels, and by the U-Mann–Whitney test ($p < 0.01$). The U-Mann–Whitney test was chosen because of the lack of normal distribution (demonstrated by Shaphiro-Wilk and Lillliefors tests). Additionally, when comparing cancer patients with controls, the percent of cancer patients with the metabolite level elevated more than the mean level plus two standard deviations was calculated.

2.5. Principal component analysis (PCA)

The PCA was performed within the Matlab environment (Matlab 7.0, Mathworks, Natick, MA, USA). Before the analysis the dataset (256 profiles × 19 metabolite levels) was autoscaled. The PCA was applied to check the dataset structure and assess the variability of the profiles belonging to groups of cancer patients vs. controls.

2.6. Hierarchical cluster analysis (HCA)

The HCA was performed employing STATISTICA tools. The Ward method, as the cluster method, and $1 − r$ (the Pearson correlation coefficient) distance, as the similarity measure, were applied to assess the similarity of variables.

2.7. Probabilistic partial least squares-discriminant analysis (p-PLS-DA)

The p-PLS-DA, which is based on boosting the PLS method [21], available in the Matlab environment, was used to measure the strength of the relationship between urinary nucleoside profiles and the presence of urogenital tract cancer. For that purpose, data after autoscalation was used. In the selection of a calibration set ($n = 160$ profiles; 80 cancer vs. 80 controls) and a test set ($n = 96$ profiles; 80 cancer vs. 16 controls) three sample selection algorithms, i.e. random selection (performed 10 times and the mean classification results were included for further consideration), the Kennard-Stone algorithm [22–24], were compared.

The duplex algorithm was selected as the most convenient. During p-PLS-DA model calibration the Monte-Carlo cross-validation was used. The discriminant model was established for all the variables considered and then less discriminatory variables were eliminated by the backward elimination procedure based on the absolute regression coefficients and model complexity (to get the smallest possible number of components), model parameters such as the root mean square error of cross-validation (RMSECV) and model predictability. To assess model predictability, the percent of correct classifications was calculated for the test set and for the calibration set as well as for the cancer patients set (model selectivity) and the controls set (model specificity).

2.8. k-Nearest Neighbor (kNN) analysis

The kNN analysis was done within the Matlab environment on autoscaled data as a nonlinear alternative to p-PLS-DA. The same calibration and test sets as in p-PLS-DA were used. In kNN model calibration, the leave-one-out cross-validation (LOOCV) was implemented to select the optimal number of neighbors and discriminatory variables to be included in the model. Predictability of the model was assessed by the percent of correct classifications for the calibration and the test sets.

3. Results

3.1. Study strategy

This study comprised of five consecutive steps. At first (step 1), nucleoside profiles were obtained for 256 subjects including 160 urogenital tract cancer patients and 96 non-cancer controls with reliable analytical procedure (Fig. 1). Secondly (step 2), changes in individual nucleosides in urogential cancer disease were evaluated by univariate analysis. Thirdly (step 3), profiles of 19 nucleosides from cancer and healthy controls were compared by unsupervised pattern recognition methods such as PCA and HCA. Next (step 4), nucleoside profiles were explored by supervised pattern recognition factors to assess the strength of the relationship of metabolite profiles with cancer presence. Additionally, the relationship of data analysis method factors such as disease stage, age and gender of the tested population were explored in the second, third and fourth step of data analysis. Finally (step 5), metabolite-to-metabolite relations within individual nucleoside profiles were assessed by the data analysis of nucleoside profiles normalized with the total peak area. All the mentioned steps 2–4 are described below.

An important issue in this and other studies with many biological samples and capillary electrophoretic methods is reproducibility of results. In presented strategy in step 1 areas of selected electrophoretic peaks are taken as quantitative information about nucleoside levels (Fig. 1). Electrophoretic peaks included in analysis were selected on the basis of results of validation studies where a set of electrophoretic peaks with acceptable reproducibility of migration times and peak areas (RSD <15%) were determined. An alternative pretreatment of electrophoretic profiles will be denosing and alignment. Afterwards whole nucleoside electrophoretic profiles could be used in further data analysis as described previously [25].

3.2. Univariate data analysis of urinary nucleoside profiles

First, individual differences in the profiles' metabolite levels between the two groups were evaluated by a comparison of statistical parameters and after performing the U-Mann–Whitney test (Table 1). The mean and median levels of each of the 19 metabolites studied were elevated in the urogenital tract cancer group in comparison to the control group. The observed differences between the
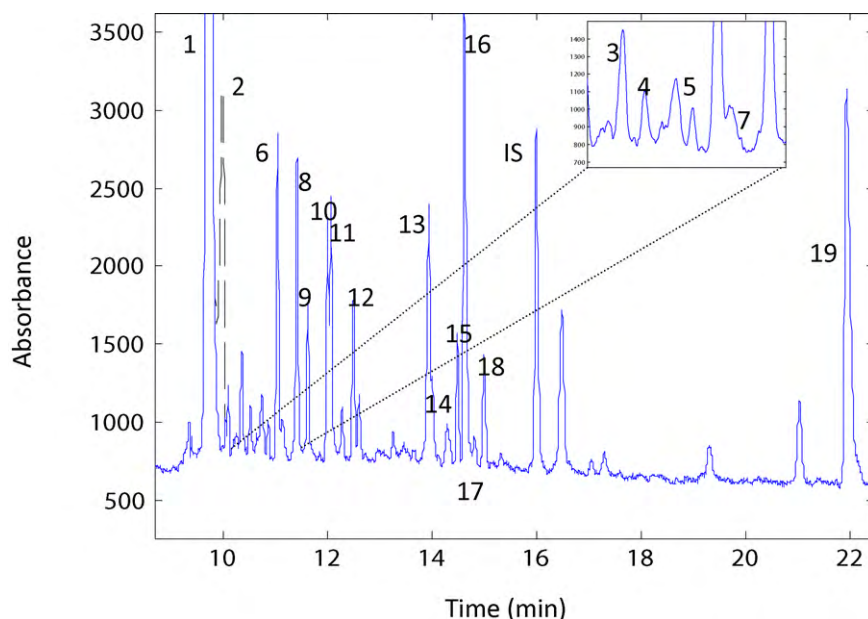
Fig. 1. Single result of the capillary electrophoretic analysis of a urinary nucleoside profile from a urine extract. Peaks: 1 – pseudouridine (pU), 2 – dihydrouridine (*dhU, measured at 214 nm), 3 – uridine (U), 4 – cytidine (C), 5 – 5-methyluridine (5mU), 6 – unidentified metabolite *1 (*1), 7 – inosine (I), 8 – unidentified metabolite *2 (*2), 9 – N4-acetylcytidine (acC), 10 – guanosine (G), 11 – unidentified metabolite *3 (*3), 12 – adenosine (A), 13 – unidentified metabolite *4 (*4), 14 – unidentified metabolite *5 (*5), 15 – unidentified metabolite *6 (*6), 16 – N2,N2-dimethylguanosine (dmG), 17 – 6-methyladenosine (6mA), 18 – xanthosine (X), 19 – 1-methyladenosine (1mA). Capillary electrophoretic conditions are: 100 mM borate, 72.5 mM phosphate, 160 mM SDS, pH 6.7; 25 kV voltage, 30 °C temperature during analysis; injection 5 s × 0.5 psi; capillary: untreated fused silica 70 cm length to detector, 50 μm I.D.

groups are statistically significant ($p < 0.01$) for most metabolites (excluding cytidine and adenosine). For 13 compounds, alternations are significant at the p-level less than 0.0001. The fold change of the group mean values is apparent and is between 1.5 and 2.0 for each metabolite (Fig. 2).

Also, the ranges of metabolite levels differ between the cancer and non-cancer groups. Ranges obtained for the cancer group are wider and comprise the ranges for the non-cancer group, which is due to dissimilar group diversity. The greater diversity of cancer group nucleoside profiles is probably the impact of some additional factors, like different type, stage and malignancy of the cancer diseases.

Next, on the basis of mean values obtained for the non-cancer group, the cutoff level could be set (cutoff level = mean + two stan-

dard deviations for each metabolite) [6,26], and the percent of cancer patients with the metabolite level above this cutoff level is calculated and categorized as elevated (Table 1). An elevated level of all the compounds was observed at *ca.* 20% of cancer subjects, varying individually between metabolites from 12.5% for adenosine to 33.75% for inosine. Therefore, none of the established metabolite cutoff levels could be alone used as an upper non-cancer level to discriminate cancer from non-cancer patients. This observation underlies the necessity to measure more than one nucleoside to evaluate the status of a subject and is in agreement with previous studies undertaken for different types of malignant diseases [6,12,26].

The dependence of nucleoside profiles on such factors like age and gender was also examined (Table 2). No statistically signif-

Table 1
Individual nucleoside/creatinine ratios [μmol nucleoside/mM creatinine] in the urine of urogenital cancer patients and non-cancer controls.

| Metabolite	Non-cancer controls ($n = 96$)			Urogenital cancer patients ($n = 160$)			Elevated level (%)	p-value	Significance
	Range	Mean ± S.D.	Median	Range	Mean ± S.D.	Median			
pU	13.67–81.69	37.18 ± 12.82	34.07	2.87–336.85	60.12 ± 47.44	44.42	23.75	0.0000	<0.0001
U	0.23–1.31	0.65 ± 0.22	0.62	0.35–7.87	1.17 ± 0.98	0.91	33.12	0.0000	<0.0001
C	0.03–0.79	0.23 ± 0.16	0.20	0.03–9.69	0.37 ± 0.81	0.23	13.75	0.1315	ns
5mU	0.20–1.67	0.74 ± 0.26	0.71	0.27–7.50	1.23 ± 1.08	0.94	25.62	0.0000	<0.0001
*1	0.69–6.47	2.55 ± 1.09	2.38	1.15–31.20	4.30 ± 3.88	3.20	25.00	0.0000	<0.0001
I	0.09–0.99	0.31 ± 0.16	0.29	0.13–4.35	0.59 ± 0.51	0.44	33.75	0.0000	<0.0001
*2	0.48–10.40	1.57 ± 1.04	1.39	0.86–16.71	2.67 ± 2.22	1.91	16.87	0.0000	<0.0001
N4aC	0.15–1.89	0.84 ± 0.33	0.76	0.05–7.37	1.27 ± 0.98	0.99	25.00	0.0001	<0.0001
G	0.09–2.93	1.06 ± 0.46	0.95	0.27–7.94	1.49 ± 1.21	1.10	18.13	0.0098	<0.01
*3	0.31–4.49	0.83 ± 0.48	0.70	0.43–8.81	1.44 ± 1.19	1.06	20.00	0.0000	<0.0001
A	0.06–1.14	0.46 ± 0.21	0.43	0.06–2.94	0.57 ± 0.44	0.45	12.50	0.2659	ns
*4	0.65–3.44	1.54 ± 0.55	1.45	0.76–11.67	2.51 ± 1.83	1.89	25.62	0.0000	<0.0001
*5	0.01–0.95	0.32 ± 0.16	0.29	0.01–3.06	0.48 ± 0.44	0.37	17.50	0.0004	<0.0005
*6	0.18–2.33	0.62 ± 0.28	0.54	0.04–4.90	0.96 ± 0.73	0.75	21.87	0.0000	<0.0001
dmG	0.93–5.70	2.54 ± 0.98	2.29	0.58–23.22	4.10 ± 3.28	3.01	19.38	0.0000	<0.0001
6mA	0.08–2.06	0.41 ± 0.24	0.39	0.04–8.68	0.68 ± 0.84	0.46	18.75	0.0008	<0.001
X	0.48–6.31	1.35 ± 0.80	1.17	0.24–10.50	2.13 ± 1.56	1.59	17.50	0.0000	<0.0001
1mA	0.65–4.91	2.14 ± 0.78	1.95	0.78–45.78	3.37 ± 4.23	2.38	19.38	0.0002	<0.0005
*dhU	2.36–30.64	9.02 ± 3.78	8.12	4.14–71.17	14.07 ± 11.05	10.36	20.00	0.0000	<0.0001

ns = no significance.
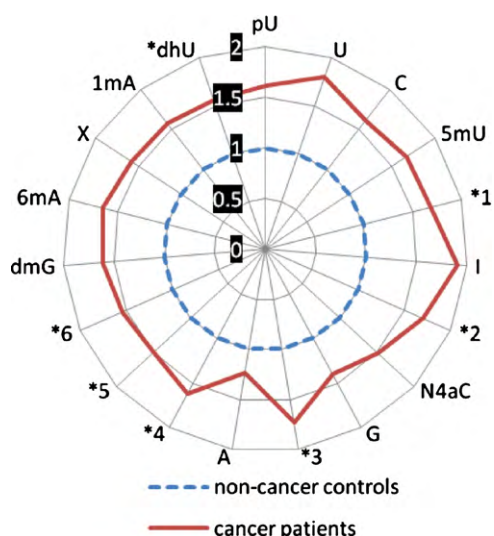
Fig. 2. Radar plot of cancer and non-cancer mean values drawn based on each mean level of each metabolite normalized to the corresponding metabolite in the non-cancer group (metabolite symbols as in Fig. 1).

icant difference at $p < 0.01$ is observed between two age groups selected from the non-cancer group. However, lower levels of all the metabolites were noticed in men in comparison to women. Inter-gender levels changes are statistically significant ($p < 0.01$) for 14 nucleosides. Changes in nucleoside levels are probably connected with inter-gender hormonal changes and *ca.* 1.2 times higher urinary creatinine day-extraction in men than in women – a result of different body muscle content [27]. Other factor as body mass index (BMI) was also included in our analysis where nucleosides profiles of non-cancer subjects with BMI lower and higher than 25 kg/m^2 were compared. No statistically relevant changes in nucleosides levels were obtained (data not shown).

3.3. Unsupervised data analysis of urinary nucleoside profiles

Before principal component analysis (PCA) metabolite levels were autoscaled to remove the dependence of the rank of the metabolites on the average concentration and the magnitude of the changes (Section 2). Results of PCA support observations pro-

vided by univariate data analysis (Fig. 3a and b). There is a fold change between profiles belonging to different groups and this change is generally reflected by the first principal component (PC1) with nearly 75% of data variance explained. At PC1/PC2 and PC1/PC2 score plots (Fig. 3a and b, respectively) the distribution of profiles from the cancer and the non-cancer groups is consistent with the diversity of these profiles observed in the analysis of individual metabolite ranges. The group of samples from non-cancer subjects is more homogeneous than group of cancer subjects and they over-lap in some extent. All the metabolites fold change has a positive impact on PC1 (Fig. 3c) and that is a main source in variation in our data. Nevertheless, the distribution of profiles in the directions of PC2 (7.3% of data variance explained) and PC3 (3.9% of data variance explained) is more dependent on the levels of individual metabolites. PC2 is affected mostly by metabolites: cytidine, 6-methyladenosine, N4-acetylcytidine, inosine and an unidentified metabolite *5, and PC3 is influenced by 1-methyladenosine. Interestingly, according to the Human Metabolome Database (HMBD) [28] and Schram [29], 1-methyladenosine and its unstable iso-mer 6-methyladenosine could be connected with urogenital tract diseases, such as chronic renal failure and different urogenital malignancies.

Next, metabolite-metabolite relationships were investigated by means of hierarchical cluster analysis (HCA) with a 1 minus cor-relation coefficient ($1 - r$) as a distance measure (see Section 2). Results of HCA provide three main groups of metabolites: A – ino-sine, 6-methyladenosine and cytidine, B – 1-methyladenosine and an unidentified metabolite *5 and C – the remaining 14 metabolites included in the profile (Supplementary material (Fig. S1)). Levels of the metabolites of group C are strongly positively intercorrelated. Hence, it could be concluded that their changes are highly redun-dant and are mainly related to the fold change of profiles between groups of non-cancer and cancer patients. On the contrary, metabo-lites from groups A and B are not so interdependent and their level changes cannot be fully explained by the fold change of metabolites from group C. This is consistent with PCA results (Fig. 3c), the more so that a similar data transformation was done in HCA and in PCA.

3.4. Supervised data analysis of urinary nucleoside profiles

The strength of the relationship between the urinary nucleoside profiles and the presence of cancer was determined by supervised

Table 2
Individual nucleoside/creatinine ratios [μmol nucleoside/mM creatinine] in urine of non-cancer controls in groups according to patient age and gender.

Metabolite	Non-cancer controls (n = 96)				Non-cancer controls (n = 96)			
	>40 years old (n = 64) Mean ± S.D.	≤40 years old (n = 32) Mean ± S.D.	p-value	Significance	Women (n = 60) Mean ± S.D.	Men (n = 36) Mean ± S.D.	p-value	Significance
pU	36.18 ± 12.11	39.20 ± 14.13	0.3469	ns	40.59 ± 12.21	31.52 ± 11.93	0.0000	<0.0001
U	0.63 ± 0.20	0.70 ± 0.26	0.3509	ns	0.70 ± 0.22	0.58 ± 0.20	0.0060	<0.01
C	0.23 ± 0.15	0.24 ± 0.16	0.7915	ns	0.27 ± 0.17	0.18 ± 0.12	0.0139	<0.05
5mU	0.71 ± 0.24	0.81 ± 0.30	0.1762	ns	0.78 ± 0.24	0.69 ± 0.29	0.0959	ns
*1	2.46 ± 0.97	2.71 ± 1.30	0.5599	ns	2.86 ± 1.03	2.03 ± 1.00	0.0000	<0.0001
I	0.29 ± 0.16	0.35 ± 0.17	0.0872	ns	0.36 ± 0.15	0.24 ± 0.17	0.0000	<0.0001
*2	1.42 ± 0.46	1.87 ± 1.67	0.1943	ns	1.75 ± 1.24	1.28 ± 0.51	0.0011	<0.005
N4aC	0.78 ± 0.29	0.94 ± 0.37	0.0278	<0.05	0.91 ± 0.33	0.71 ± 0.28	0.0011	<0.005
G	0.83 ± 0.54	0.83 ± 0.34	0.5289	ns	0.93 ± 0.55	0.66 ± 0.26	0.0000	<0.0001
*3	1.07 ± 0.49	1.05 ± 0.40	0.9256	ns	1.18 ± 0.46	0.87 ± 0.41	0.0003	<0.0005
A	0.44 ± 0.19	0.50 ± 0.24	0.3672	ns	0.52 ± 0.21	0.35 ± 0.16	0.0000	<0.0001
*4	1.51 ± 0.53	1.59 ± 0.61	0.7033	ns	1.63 ± 0.58	1.38 ± 0.47	0.0227	<0.05
*5	0.29 ± 0.12	0.38 ± 0.20	0.0345	ns	0.34 ± 0.16	0.28 ± 0.15	0.0201	<0.05
*6	0.64 ± 0.30	0.59 ± 0.24	0.4555	ns	0.67 ± 0.30	0.55 ± 0.24	0.0109	<0.05
dmG	2.48 ± 0.95	2.64 ± 1.04	0.4416	ns	2.79 ± 0.96	2.11 ± 0.85	0.0000	<0.0001
6mA	0.40 ± 0.26	0.43 ± 0.19	0.2194	ns	0.45 ± 0.28	0.34 ± 0.16	0.0045	<0.005
X	1.22 ± 0.57	1.59 ± 1.09	0.0803	ns	1.51 ± 0.87	1.07 ± 0.56	0.0000	<0.0001
1mA	2.08 ± 0.77	2.25 ± 0.80	0.3430	ns	2.39 ± 0.74	1.72 ± 0.66	0.0000	<0.0001
*dhU	8.85 ± 4.03	9.37 ± 3.25	0.3123	ns	9.85 ± 3.94	7.64 ± 3.08	0.0003	<0.0005

ns = no significance.

Fig. 3. PCA results. (a and b) PC1/PC2 and PC1/PC3 scoreplots (● Non-cancer controls, ✳ urogenital tract cancer patients), (c) PC1/PC2/PC3 biplot (metabolite symbols as in Fig. 1).

Table 3
Overview of diagnostic outcomes of supervised pattern recognition analysis: p-PLS-DA and kNN.

Parameter/model	p-PLS-DA	kNN
LVs/k (n)	4 (19)	3 (4)
CCR_cal	81.25	69.37
CCR_cal_c	80	91.25
CCR_cal_h	82.5	47.5
CCR_test	71.87	80.21
CCR_test_c	72.5	88.75
CCR_test_h	68.75	37.5
CCR_total	77.73	73.44
SE (selectivity)	76.25	90
SP (specificity)	80.21	45.83

Models symbols: LVs = number of latent factors, k = number of neighbors, n = number of metabolites included in the model, CCR = correct classification rate, cal = calibration set, test = test set, total = all dataset, c = cancer group, h = non-cancer group.

pattern recognition methods like the nonlinear and nonparametric k-Nearest Neighbor (kNN) method and the method of linear and parametric partial least squares-discriminant analysis with a probabilistic function (p-PLS-DA) [21]. The results of supervised methods are more powerful than results of exploratory, unsupervised data analysis tools as PCA (Fig. 3) because they include information about class membership.

Before supervised pattern recognition analysis, the dataset was divided into a calibration and a test set with the help of one of three selection methods: random selection, the Kennard-Stone algorithm [22,23] and the duplex algorithm [24] (see Section 2). The duplex method, which guarantees the representativeness of both subsets (calibration and test sets) was the most convenient in the case of our data and was selected further. In Table 3 the diagnostic outcomes of optimized p-PLS-DA and kNN models are shown. Correct classification rates obtained in kNN analysis are in general lower than those provided by p-PLS-DA. Worse kNN results are probably due to the different diversity of groups and their partial overlapping in multidimensional space (Fig. 3). The best p-PLS-DA model (model A3 in Table 3) provides 77.3% of properly classified samples (198 of a total of 256 samples) with 76.5% sensitivity and 80.2% specificity. That is a major improvement in comparison to the results of the analysis of elevated levels of individual metabolites (Table 1). Moreover, it supports the hypothesis behind the metabolomic approach that a group of potential biomarkers is more likely to form patterns allowing for disease recognition than any single compound is able to do, be it identified or not.

Next when the misclassified samples from the cancer group were studied, no relation between the type or severity of disease and the percent of misclassified patients could be further found. Also, no dependence of model predictability on patient gender could be identified (Table 4), in contrary to the gender dependency found for individual metabolites (Table 2). Also, when p-PLS-DA models were developed separately for male and female study participants, no evident improvement in classification rates of new models was observed (Table 4).

3.5. Analysis of metabolite-metabolite relations within individual nucleoside profiles

In previous studies [6,12,26] the relation between the fold change of individual urinary nucleosides and the appearance of cancer was set. Results of our large-scale metabolomic study also showed this relation in urogenital cancer diseases. Therefore, an important question arises – what are the characteristics of alterations in urinary nucleoside profiles between the urogenital tract cancer group and non-cancer group and how do these alterations impact the discrimination of the two studied groups of patients? In

Table 4
Overview of supervised pattern recognition (p-PLS-DA) results.

Parameter/model	F3	M3	N3
LVs (n)	2 (19)	3 (19)	1 (19)
CCR_cal	85	86.67	80
CCR_cal_c	83.33	76.67	70
CCR_cal_h	86.67	96.67	90
CCR_test	80	44.79	48.95
CCR_test_c	100	44.22	46.25
CCR_test_h	73.33	83.33	62.5
CCR_total	83	60.89	68.37
SE (selectivity)	87.5	50.8	58.13
SP (specificity)	80	94.4	85.42

Symbols: F – model obtained for women (156 profiles), M – model obtained for men (100 profiles), N – model obtained for data normalized to total peak area (256 profiles), calibration set and test set selected with duplex algorithm (3). LVs = number of latent factors, k = number of neighbours, n = number of metabolites included in the model. CCR = correct classification rate, cal = calibration set, test = test set, total = all dataset, c = cancer group, h = non-cancer group.

other words: is a change of urinary nucleoside profiles in urogenital cancer diseases only a quantitative level change of all metabolites or also a change of level ratios of metabolites?

It could be concluded that nucleoside-to-nucleoside relations could change in urogenital cancer diseases because not all the metabolites are strongly positively intercorrelated (see HCA results in Supplementary material (Fig. S1)). To establish what character the relationships have between metabolite-metabolite relations and the appearance of urogenital cancer, differently preprocessed urinary nucleoside profiles were evaluated. These profiles were not creatinine normalized but were normalized to the total peak area to overcome the differences in the concentrations between the samples (neglecting profiles' fold change) [30,31]. In this way, the main focus was set on the composition of urinary nucleoside profiles aside from the magnitude of their urinary levels (urinary offset level). When differences in the composition of profiles between the cancer and non-cancer groups were analyzed, significant alterations ($p < 0.01$) in the content of nine metabolites were observed. These changes were in both directions, e.g. the inosine relative content was higher for cancer patients whereas the 1-methyladenosine content was higher for non-cancer controls. Nevertheless, these alterations were not sufficient to successfully discriminate urogenital cancer from non-cancer patients and to build a discriminant model with similar predictability as the previously obtained model A3 (Table 4).

4. Discussion

It has been well established that the levels of some urinary nucleosides could be changed in malignant diseases, such as leukemia or liver cancer [6,12,26,27]. However, it was still not clear how the much more complex urinary nucleoside profiles alterate in urogenital cancer diseases and what the character and strength of these alterations is. The results presented here demonstrate that changes of complex urinary nucleoside profiles are dependent upon the presence of urogenital cancer and that these changes are both qualitative and quantitative in nature. The metabolomic approach applied in exploring urinary nucleoside profiles allowed a validation of the relationship between these metabolite profiles and urogenital cancer diseases and provided a good prediction model for these disorders. In forthcoming studies it should be investigated whether 76.5% sensitivity and 80.2% specificity provided by the optimized discriminant model of the studied urinary nucleoside profiles might be successfully employ in the diagnosis of urogenital cancer diseases. The answer to this will be positive when a high diversity of cases included and hence the possible method universality will be considered. The limited number and diagnos-

tic power of currently used diagnostic tools in urogenital cancer diseases [32,33] could be overcome by the metabolomic approach and nucleoside profiles. The presented approach has already been submitted to obtain a patent. Hopefully, the presented study is a step forward to finally apply the dependency of urinary nucleoside profiles to urogenital tract cancer detection in clinical practice. Nevertheless, larger groups of patients and more precise information on factors which may affect the urinary nucleoside profiles (e.g. coexisting diseases or applied cancer therapy) have to be included in further research to fully validate the postulated relationship between the urinary nucleoside profiles and the appearance of urogenital tract cancer.

In this paper, a study of urinary nucleoside profiles by the metabolomic approach including the analytical procedure specific to nucleosides and the extensive data analysis of collected metabolite profiles is comprehensively described. The used approach provided new essential information about the relations between changes of urinary nucleoside profiles and the presence of urogenital cancer diseases which could be employed in the discrimination of cancer subjects from healthy controls.

Acknowledgments

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jpba.2010.07.031.

References

[1] L.K. Schnackenberg, Global metabolic profiling and its role in system biology to advanced personalized medicine in the 21st century, Expert Rev. Mol. Diagn. 7 (2007) 247–259.
[2] A. Nordstrom, R. Lewensohn, Metabolomics: moving to the clinic, J. Neuroimmun. Pharmacol. 5 (2010) 4–17.
[3] J.L. Spratlin, N.J. Serkova, S.G. Eckhardt, Clinical applications of metabolomics in oncology: a review, Clin. Cancer Res. 15 (2009) 431–440.
[4] Y. Ma, G. Liu, M. Du, I. Stayton, Recent developments in the determination of urinary cancer biomarkers by capillary electrophoresis, Electrophoresis 25 (2004) 1473–1484.
[5] B. Melichar, D. Solichova, I. Svobodova, K. Melicharova, Neopterin in renal cell carcinoma: inhalational administration of interleukin-2 is not accompanied by a rise of urinary neopterin, Luminescence 20 (2005) 310–314.
[6] A. Seidel, S. Brunner, P. Seidel, G.I. Fritzm, O. Herbarth, Modified nucleosides: an accurate tumour marker for clinical diagnosis of cancer, early detection and therapy control, Br. J. Cancer 94 (2006) 1726–1733.
[7] J.A. Byun, S.H. Lee, B.H. Jung, M.H. Choi, M.H. Moon, B.C. Chung, Analysis of polyamines as carbamoyl derivatives in urine and serum by liquid chromatography-tandem mass spectrometry, Biomed. Chromatogr. 22 (2008) 73–80.
[8] K.R. Kim, S. La, A. Kim, J.H. Kim, H.M. Liebich, Capillary electrophoretic profiling and pattern recognition analysis of urinary nucleosides from uterine myoma and cervical cancer patients, J. Chromatogr. B 754 (2001) 97–106.
[9] Y. Masaki, K. Itoh, T. Sawaki, H. Karasawa, T. Kawanami, T. Fukushima, H. Kawabata, Y. Wano, Y. Hirose, T. Suzuki, S. Sugai, H. Umehara, Urinary pseudouridine in patients with lymphoma: comparison with other clinical parameters, Clin. Chim. Acta 371 (2006) 148–151.
[10] J. Yang, G. Xu, Y. Zheng, H. Kong, T. Pang, S. Lv, O. Yang, Diagnosis of liver cancer using HPLC-based metabonomics avoiding false-positive result from hepatitis and hepatocirrhosis diseases, J. Chromatogr. B 813 (2004) 59–65.
[11] Y.F. Zheng, H.W. Kong, J.H. Xiong, S. Lv, G.W. Xu, Clinical significance and prognostic value of urinary nucleosides in breast cancer patients, Clin. Biochem. 38 (2005) 4–30.

[12] G. Xu, H.M. Liebich, Normal and modified nucleosides in urine as potential tumor markers determined by MEKC and HPLC, Am. Clin. Lab. 3 (2001) 22–32.

[13] C.W. Gehrke, K.C. Kuo, G.E. Davis, R.D. Suits, T.P. Waalkes, E. Borek, Quantitative high-performance liquid chromatography of nucleosides in biological materials, J. Chromatogr. 150 (1978) 455–476.

[14] Y. Zheng, G. Xu, J. Yang, X. Zhao, T. Pang, H. Kong, Determination of urinary nucleosides by direct injection and coupled-column high-performance liquid chromatography, J. Chromatogr. B 819 (2005) 85–90.

[15] S.-H. Cho, B.H. Jung, S.H. Lee, W.-Y. Lee, G. Kong, B.C. Chung, Direct determination of nucleosides in the urine of patients with breast cancer using column-switching liquid chromatography-tandem mass spectrometry, Biomed. Chromatogr. 20 (2006) 1229–1236.

[16] K.-R. Kim, S. La, A. Kim, J.-H. Kim, H.M. Liebich, Capillary electrophoretic profiling and pattern recognition analysis of urinary nucleosides from uterine myoma and cervical cancer patients, J. Chromatogr. B 754 (2001) 97–106.

[17] Y.-X. Zhang, Artificial neural networks based on principal component analysis input selection for clinical pattern recognition analysis, Talanta 73 (2007) 68–75.

[18] T. Helboe, S.H. Hansen, Separation of nucleosides using capillary electrochromatography, J. Chromatogr. A 836 (1999) 315–324.

[19] M.J. Markuszewski, W. Struck, M. Waszczuk-Jankowska, R. Kaliszan, Metabolomic approach for determination of urinary nucleosides as potential tumor markers using electromigration techniques, Electrophoresis 31 (2010) 2300–2310.

[20] E. Szymańska, M.J. Markuszewski, K. Bodzioch, R. Kaliszan, Development and validation of urinary nucleosides and creatinine assay by capillary electrophoresis with solid phase extraction, J. Pharm. Biomed. Anal. 44 (2007) 1118–1126.

[21] M.H. Zhang, Q.S. Xu, D.L. Massart, Boosting partial least squares, Anal. Chem. 77 (2005) 1423–1431.

[22] R.W. Kennard, L.A. Stone, Computer-aided design of experiments, Technometrics 11 (1969) 137–148.

[23] I. Stanimirova, A. Kubik, B. Walczak, J.W. Einax, Discrimination of biofilm samples using pattern recognition techniques, Anal. Bioanal. Chem. 390 (2008) 1273–1282.

[24] R.D. Snee, Validation of regression models. Methods and examples, Technometrics 19 (1977) 415–428.

[25] E. Szymańska, M.J. Markuszewski, X. Capron, A.-M. van Nederkassel, Y. Vander Heyden, M. Markuszewski, K. Krajka, R. Kaliszan, Increasing conclusiveness of metabonomic studies by cheminformatic preprocessing of capillary electrophoretic data on urinary nucleoside profiles, J. Pharm. Biomed. Anal. 43 (2007) 413–420.

[26] H.M. Liebich, G. Xu, C. Di Stefano, R. Lehmann, Capillary electrophoresis of urinary normal and modified nucleosides of cancer patients, J. Chromatogr. A 793 (1998) 341–347.

[27] C.M. Slupsky, K.N. Rankin, J. Wagner, H. Fu, D. Chang, A.M. Weljie, E.J. Saude, B. Lix, D.J. Adamko, T.J. Marrie, Investigations of the effects of gender, diurnal variation, and age in human urinary metabolic profiles, Anal. Chem. 79 (2007) 6995–7003.

[28] D.S. Wishart, et al., HMDB: the human metabolome database, Nucleic Acids Res. 35 (2007) D521–D526.

[29] K.H. Schram, Urinary nucleosides, Mass Spectrom. Rev. 17 (1998) 131–251.

[30] T. Kind, V. Tolstikov, O. Fiehn, R.H. Weiss, A comprehensive urinary metabolic approach for identifying kidney cancer, Anal. Biochem. 363 (2007) 185–195.

[31] M. Daszykowski, W. Wu, A.W. Nicholls, R.J. Ball, T. Czekaj, B. Walczak, Identifying potential biomarkers in LC–MS data, J. Chemometr. 21 (2007) 292–302.

[32] S. Sozen, S. Eskicorabci, H. Ozen, Urinary markers for urothelial cancer, Br. J. Urol. 92 (2003) 531–533.

[33] E. Schiffer, Biomarkers for prostate cancer, World J. Urol. 25 (2007) 557–562.